DOCUMENT RESUME

ED 114 825 CS 202 337

AUTHOR	White, Edward M.
TITLE	Comparison and Contrast: The 1973 California State University and Colleges Freshman English Equivalency Examination.
INSTITUTION	English Council of the California State Univ. and Colleges.
PUB DATE	Oct 73
NOTE	183p.; See related document CS 202 338
AVAILABLE FROM	Office of the Chancellor, the California State University and Colleges, 5670 Wilshire Blvd., Los Angeles, California 90036 ($1.50, make checks payable to The State of California)
EDRS PRICE	MF-$0.76 HC-$9.51 Plus Postage
DESCRIPTORS	Achievement Tests; *College Freshmen; *Comparative Analysis; *Composition (Literary); Educational Research; *Equivalency Tests; Higher Education; *Literature Appreciation; Measurement Instruments; Test Construction; Test Validity
IDENTIFIERS	California; *College Level Examination Program

ABSTRACT

In the late spring of 1972, the Chancellor's Office agreed to support a summer study to be undertaken by a committee of the California English Council, to investigate equivalency testing in the area of English and to recommend an appropriate program for use by the California State University and Colleges. This report is the result of that study: it attempts to focus on the major issues in such a way as to point to their solution, and it recommends a method of equivalency testing in English which is responsive to the discipline and practical to implement. The English examination administered in May of 1973 consisted of a 90-minute essay test conducted and graded by California State University and College English faculty, and the 90-minute objective CLEP Subject Examination, "Analysis and Interpretation of Literature," which was also administered at that time. The most significant results of this study were that 1,362 out of 4,071 students taking the test were identified as deserving college credit for freshman English, and the essay component has been shown to be a valuable part of freshman English equivalency testing. (The study and its results are described in both narrative and table format with newspaper clippings and news releases included.) (RB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED114825

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

COMPARISON AND CONTRAST

THE 1973 CALIFORNIA STATE UNIVERSITY AND COLLEGES FRESHMAN ENGLISH EQUIVALENCY EXAMINATION

EDWARD M. WHITE
DIRECTOR OF THE PROJECT
FOR
THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

A limited number of copies of this document are available for purchase from the Office of the Chancellor, the California State University and Colleges, for $2.00 per copy. Checks or money orders should be made payable to The State of California. Add 6% sales tax for California addresses. In addition, ERIC compatible microfiche copies are available on some reports published after January 1, 1972.

This monograph is one in a series describing the results of projects implemented under The California State University and Colleges Fund for Innovation and Improvement in the Instructional Process.

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES ENGLISH EQUIVALENCY EXAMINATION

1973

By

Edward M. White

Director of the project for the
California State University and Colleges
English Council

Project 72-143
Fund for Innovation and Improvement

October 1973

TABLE OF CONTENTS

I. Summary and Fact Sheet 1

II. Administration of the Pilot Project 14

A. Personnel 14

1. Project Directors 14

2. Coordinator of Test Administration 15

3. Statistician 16

4. Essay Question Leaders 17

5. Aides 18

B. Meetings and Decisions 19

1. Single Closed Administration 21

2. Location of Test Centers 21

3. Day of Test 22

4. Date of Test 23

5. Time of Day 23

6. Proctors 24

7. Payment of Proctors 24

8. Movement of Test Materials 25

9. Closing Date and Late Registration 26

10. Test Manual 26

11. Test Form 27

12. Registration Form 27

C. The Objective Test 28

D. The Essay Test 30

1. Format 30

2. Creation and Aims of the Essay 31

3. Grading System 32

4. Question 1 33

a. Directions to Students. 34

b. Directions to Graders 34

c. Report by Question Leader: Dean Gerhard Friedrich. 35

Six Sample Student Papers 37

5. Question 2 42

a. Directions to Students. 42

b. Directions to Graders 42

c. Report by Question Leader: Dr. Rex Burbank 44

d. Six Sample Student Papers 47

6. The Norm Sample. 51

7. The Essay Reading. 58

E. Converting the Essay Grade and Combining Scores. 62

F. Determination of the Passing Grade. 65

G. Notification of Results 68

H. Public Information. 71

IV. Follow-up Studies. 74

V. The 1974 English Equivalency Examination. 77

VI. Charts and Tables. 79

A. Distribution of Composite Scores - Total Applicant Population. 80

B. Distribution of Composite Scores - Applicants Who Passed 81

C. Distribution of Composite Scores - Applicants Who Did Not Pass 82

D. Distribution of Converted Essay Scores - Total Applicant Population. 83

E. Distribution of Converted Essay Scores - Applicants who Passed 84

F. Distribution of Converted Essay Scores - Applicants who did not Pass 85

G. Distribution of Objective Scores - Total Applicant Population 86

H. Distribution of Objective Scores - Applicants who Passed 87

I. Distribution of Objective Scores - Applicants who did not Pass 88

J. Intercorrelations Among Part and Composite Scores - Total Applicant Population 89

K. Intercorrelations Among Grouped Essay and Composite Scores - Total Applicant Population 90

L. Male Applicants by Age Category 91

M. Female Applicants by Age Category 92

VII. Appendix

A. Equivalency Testing in College Freshman English 94

B. Budget for the 1974 English Equivalency Test 126

C. Statistician's Reports: Dr. William Cowell. 132

1. Statistical Procedures,May, 1973 132

2. Statistical Report, July, 1973 150

3. Statistical Report, February, 1974 159

D. Newspaper Reactions to the Project. 168

I

SUMMARY AND FACT SHEET

This chapter, with the appended fact sheet, summarizes the results of the Freshman English Equivalency Examination that was administered to 4,071 students on May 12, 1973. Analysis of the results shows that this program, designed and administered by the California State University and Colleges English Council, and funded by the Chancellor's Office division of New Program Development, was a major success; it provided clear and substantial benefits to the 1,362 students who earned six semester units of college credit, to the English faculties, and to the California State University and College system. Furthermore, the project embodied a constructive resolution to what had seemed to be an irreversible and bitter conflict.

For many years, English faculties have been largely opposed to the practice of large-scale equivalency testing and to the objective testing instruments generally used. This opposition was not without a basis in fact: most objective tests of writing ability, in particular, do not measure what is usually taught in freshman English.

English teachers, then, have generally opposed external testing programs. These faculty have not had the funds, the time, or the specialized statistical knowledge

that are essential to develop large scale programs of their own. Thus, it is not surprising that administrators, generally eager to proceed with equivalency testing, have been perplexed at the positions English departments have taken in this area. A particularly vivid example of this faculty-administration misunderstanding occurred in the fall of 1971, when large-scale equivalency testing was in fact instituted on two California State University and College campuses, arousing substantial opposition from English departments.[1]

The California State University and Colleges English Council took a statesmanlike position in the fall of 1971, rejecting what it called "an improper objective test," but endorsing "the principle of properly constructed and properly administered challenge examinations." Much constructive discussion and committee work followed.[2] By the summer of 1972, the Chancellor's Office and the

---

[1] For a convenient summary of the results of this testing program, see Urban Whitaker, "Credit by Examination at San Francisco State," College Board Review, 83 (Spring, 1972), 12-16. The conflict at its most heated appears in an exchange between Whitaker and Vernon T. Hornback, then president of the California State University and Colleges English Council, in The Future of General Education in the California State University and Colleges, Office of the Chancellor (1972), 59-76.

[2] A committee on equivalency testing established in the Chancellor's Office was under the chairmanship of Dr. Gerhard Friedrich; its sub-committee on English was under the chairmanship of Professor Charles Adams, of California State University, Chico. Professor James Clark, California State University, San Jose, was chairman of the English Council committee. The statewide Academic

Educational Testing Service had combined resources to support an English Council report and proposal, calling for a professionally acceptable examination.[3]

The success of the Spring 1973 test has implications that extend outside the field of English and beyond the borders of California; it has been shown that faculty, administrators, and national testing agencies can work together in harmony for the benefit of students and the whole educational process, even when mistakes have been made that have produced an emotion-charged reaction. The project has also shown that properly constituted faculty leadership, when given adequate support, is an appropriate means of achieving academic innovation.

The examination administered May 12, 1973 consisted of a 90-minute essay test constructed and graded by California State University and College English faculty, and the 90-minute objective CLEP Subject Examination, Analysis and Interpretation of Literature.[4] That objective test was the only such test to gain the endorsement of the California State University and Colleges English Council,

---

Senate, through its Educational Policies Committee, chaired by Jerome Fox, California State University, San Francisco, also took an active interest in the matter.

[3] Edward M. White, Equivalency Testing in College Freshman English: A Report and a Proposal, 1972. Available through ERIC. (See Appendix, page 94)

[4] The College-Level Examination Program (CLEP) is sponsored by the College Entrance Examination Board (CEEB) and administered by the Educational Testing Service (ETS).

after careful scrutiny of objective tests available in 1972.[5]

The results of the project are undergoing extensive analysis, particularly in the Chancellor's Office division of Institutional Research, which will be assisting the English Council in producing follow-up studies over the next several years. The report which follows contains a full account of the development of the project, including all information available by November 1, 1973. The most significant results now apparent can be summarized briefly:

1. 1,362 students deserving college credit for freshman English were identified by a rigorous, responsible, and fair test. These students will be able to enter English courses at an appropriate advanced level, with positive feelings towards themselves and their English studies, and, if they wish, to accelerate their college careers.
2. Student writing has been shown to be a valuable part of freshman English equivalency testing. Those objecting to essay testing have generally cited problems of reliable grading and high cost. This project has demonstrated that

---

[5] A new CLEP Subject Examination in Freshman English has since been released, and it has received some favorable comment. The new examination should not be confused with the CLEP General Examination in English Composition, which has been declared inappropriate by conferences of English teachers in California, Illinois, Florida, and elsewhere. A CLEP Subject Examination in English Composition also exists, but it has been described as inappropriate by the CSUC English Council. It should not be surprising if confusion results from three CLEP tests in the same area.

both those objections can be overcome. The carefully controlled essay reading conducted June 16-20, 1973, by seventy-five English professors drawn from all nineteen California State University and College campuses, produced highly accurate test scores at relatively modest cost.

The high reliability of the essay scoring was a result of careful planning. The essay questions were composed by a representative and experienced group of English professors, who carefully worked out agreements on the nature of appropriate questions and on the criteria to be used in judging the candidates' writing. The readers were selected after nomination by department chairmen, and only regular faculty with a demonstrated active concern for freshman English were nominated. The grading session schedule contained time for frequent discussions of standards, using sample papers; the readers were carefully trained to use common criteria for grading; four independent readings were given to each examination, and, where significant differences in evaluation occurred, papers were reread to reconcile the differences; finally, there were systematic reviews and cross-checks on individual ratings to ensure consistency throughout the reading.

The basic argument for essay testing, however, does not depend on statistical descriptions of test reliability, though essay testing can indeed be reliable. Almost everyone will agree that a college student should be able to write with directness, clarity, and precision; obviously,

anything motivating practice and instruction in writing serves fundamental educational goals. The existence of a valid and reliable test demanding writing, as part of a freshman English equivalency examination, is likely to have substantial and positive effects on teaching and learning in the schools as well as in the colleges.

J. Start-up funding for a similar test in 1974 has been provided by the Chancellor's Office, which has expressed the intention of exploring with the California Department of Finance ways of incorporating equivalency testing programs into the faculty workload budget. A preliminary analysis shows that the credit hours earned cost the State of California much less than the usual expense for instruction.

However, no one should assume that the low cost of these credits means that the credits were cheaply earned. It is necessary to set high test standards to make sure that students who have in fact not gained college-level abilities do not receive college credit; such students should not be deprived of necessary educational experiences by equivalency credit awarded too easily. Furthermore, it is not to be expected that a very high percentage of students without a college course should perform as well as students passing a year's work in college English. The 1973 test has shown that, despite a rigorous test, graded with high standards, one-third of the test group passed. In fact, the test group, perhaps because of the $15.00

test fee, seemed unusually able and highly motivated; those who passed out-performed (on the essay test) most of the college students who had successfully completed the college course and whose papers were graded by the same standards.[6]

As long as the academic integrity of the testing process is safeguarded by responsible faculty direction, reasonable people can only applaud the savings that equivalency testing offers to colleges and to the able and ambitious students who succeed.

4. The California State University and College system has attracted nationwide attention for its leadership in English Equivalency Testing. The policies and reports of the English Council have been widely reprinted, and the financial support of English Council proposals by the Chancellor's Office has set new precedents. The director of the project has been asked to make major presentations at major conferences in New York, Lousiana, Illinois, Washington, and Utah; conferences in Texas and Florida have accepted and endorsed parts or the whole of the English Council report; newspapers and educational journals have reported favorably on the project; and ETS and CEEB have shown an increasing disposition to follow the California model in the future.

---

[6] The norm sample of college essays turned out to be unrepresentative of the California State University and Colleges system and could not be used for setting cutting

5. Some results will be hard or impossible to measure, but are nonetheless significant. The existence of the test is likely to improve the preparation in English that secondary schools provide to their students; the test should help strengthen curriculum and encourage attention to writing skills. The existence of the essay grading session, bringing English professors from all campuses together for evaluation of student writing under specialized direction, is bound to improve aspects of college English teaching. Perhaps most important of all, the way in which equivalency testing shifts responsibility for learning onto the shoulders of students has large and positive implications for education as a whole, as any teacher looking at an essentially unmotivated class will agree.

Finally, the director of this project wishes to thank the following, whose assistance, cooperation, encouragement, and support helped make the project a success:

Members of the California State University and Colleges English Council, i.e., the department chairmen and freshman English coordinators, who gave substantial amounts of time and energy to the project; and particularly Dr. Richard Lid, California State University, Northridge, who served as co-director of the project, helping with all phases of the

---

scores. But the sample did show that students passing the test were performing better than many passing college students.

work and handling with aplomb the tangled problems of budget, facilities planning, and correspondence.

Members of the California State University and Colleges English Council committee that worked with the project directors to draw up the questions and grading criteria for the essay test: Professor Rex Burbank, California State University, San Jose; Professor Michael Cartwright, California State College, Bakersfield; Professor Will Crockett, California State University, San Jose; State University Dean Gerhard Friedrich; Professor Eileen Lothamar, California State University, Long Beach; Professor Bill Leary, California State University, Los Angeles.

Members of the Chancellor's staff, particularly Dr. David Provost, State University Dean, New Program Development, Dr. Jack Smart, Deputy Dean and Dr. David Leveille, Associate Dean in the same office; Dr. Gerhard Friedrich, State University Dean, Academic and Resource Planning; Dr. Robert Bess, Director, Academic Projects; Dr. Leon Thomas, Associate Dean, Institutional Research; and Mr. Charles Davis, Public Affairs Associate.

Test Officers on the nineteen California State University and College campuses, who undertook unusual responsibilities for this program, particularly Dr. William Abbott and Dr. Richard Cantey at California State University, Long Beach, who coordinated communications with the other test officers and assisted the program in many ways.

The staff of the Educational Testing Service, particularly Dr. William Cowell, from the Princeton office, whose statistical assistance was invaluable; Mr. Alan Seder from the Berkeley office, whose experience and tact were equally invaluable; Dr. Richard Harsh, from the Los Angeles office; Dr. Albert Serling, Program Director for the College-Level Examination Program in Princeton, and his assistant Ms. Betsy Barlow.

1973 ENGLISH EQUIVALENCY EXAMINATION FACT SHEET

Test Date:	May 12, 1973
Test Used:	CLEP Subject Examination, Analysis and Interpretation of Literature (90-minutes) in combination with Two 45-minute essay questions prepared by CSUC English faculty (90-minutes)
Number of Candidates:	4,071
Number Passed:	1,362
Percentage Passed:	33.5%
Number of Credit Hours Earned:	8,172

Scoring Data:

Minimum Passing Scores:

Objective Part:	45 (CLEP scale)(achieved by 60.3%)
Essay Part:	13 (out of a possible 24 points; equivalent to 48 on the CLEP scale.)(achieved by 53.5%)
Combined Score:	100 (49 objective + 51 essay, each on the CLEP scale.)(42.6% achieved 49 or better on the objective test, and 40.3% achieved 51 or better on the essay test.)(153 students achieved a combined score of 100 or more but did not pass because they did not achieve a minimum pass on one of the parts.)

Mean, Median, Mode, Standard Deviation:

Objective Part:

Mean:	47.448
Mode:	43.0
Median:	47.149

Standard Deviation:	8.854
Essay Part:	
Mean:	47.916
Mode:	48.0
Median:	47.662
Standard Deviation:	8.402
Combined Score:	
Mean:	95.363
Mode:	91.0
Median:	94.485
Standard Deviation:	14.869
Essay Reading:	Kellogg-West Conference Center, June 16-20, 1973
Number of Readers:	75
Number of Colleges Represented:	All 19 CSUC campuses
Grading Scale:	6-point
Number of Readings per Test:	4 (2 independent readings for each question; additional readings to reconcile discrepant grades)
Total Essay Readings:	Approximately 20,000
Weighting and Scaling:	The essay score was converted by the equipercentile method to the CLEP scale (20-80), and the scores on both parts of the examination were added. Each part received equal weight.
Statistical Data:	
Correlation between CSUC Essay and CLEP Objective Tests:	.4777

Reliability:

CLEP Test:	.88
Essay Test:	.7183 ± .1802
Composite Score:	.88 ± .05

Student Profile:

Test Population:

Female:	2,354, 97.1% age 19 or under
Male:	1,702, 97.1% age 19 or under

Pass Rate:

Female:	40.1%
Male:	24.3%

Budget Data:

Allotted by State of California (from New Program Development Funds):	$64,003.00
Test Fees at $15.00 per Student:	$61,065.00
Estimated Overhead and Miscellaneous Costs Contributed by Institutions:	
CSU Northridge:	$ 2,000.00
CSC San Bernardino:	$ 7,715.00
CLEP:	$28,000.00
Cost to State of California per Credit Hour Earned:	$ 9.00 (approximately)

III

ADMINISTRATION OF THE PILOT PROJECT

A. Personnel

1. Project Directors. The October 26, 1972 California State University and Colleges English Council meeting concluded with the appointment of Professor Edward M. White, Chairman, Department of English, California State College, San Bernardino, as Director of the Test Program for the English Council. The council also appointed Professor Richard Lid, then Chairman, Department of English, California State University, Northridge, as assistant to the Director of the project. As the project developed, Professors White and Lid became co-directors, and Professor Lid assumed responsibilities in the area of budget, facilities planning, and correspondence. Since it proved more flexible to have the program budget established at the Northridge campus, Professor Lid also assumed supervision over the project secretarial staff and the disbursement of funds. Professor White remained responsible for overall project direction, development of the test and test grading procedures, and relations with the English Council, the Chancellor's Office, and the Educational Testing Service. It proved to be a very useful division of labor for routine matters; however, on all important decisions, the project directors agreed and worked together.

2. Coordinator of Test Administration. A series of additional appointments were made as the project progressed and as the need for them became apparent. In January, it became clear that the administration of the test itself would require a substantial amount of time, training, and personal contact. Dr. William Abbott, Test Officer, California State University, Long Beach, agreed to act as coordinator of test administration, so that the nineteen separate test offices which would be administering the test on May 12 could have one central location for information and direction. Dr. Abbott and his very able assistant, Dr. Richard Cantey, relieved the project directors of an immense amount of detail work and performed a valuable function in the course of the test administration. Certainly, any administration of a test on multiple campuses requires a similar diligent test officer to coordinate the activities of the various test administrators.

Among the functions that the test office performed are the following: 1) assistance in the preparation of a memo to all test officers (April 1973) providing basic information on the test program and requesting their participation, most particularly the immediate selection of test locations on each campus; follow-up memos were sent on May 8, 1973 and June 10, 1973; 2) assistance in the preparation of the test registration

form, and communication with the various test officers on proper ways to handle that form; 3) coordination and direction of all test materials in cooperation with the Educational Testing Service office in Berkeley; 4) preparation of a test manual for all test officers and proctors, in co-operation with the Educational Testing Service, Princeton; 5) accumulation of information about the administration of the test for use by the project directors shortly after the administration; the California State University, Long Beach Test Office made approximately 200 telephone calls in connection with the project; 6) assistance in determing the statistical procedures to be used in combining scores and finding cutting scores.

3. Statistician. As the test date approached, it became necessary to make several additional administrative appointments. Happily for the project, Dr. William Cowell, formerly Statistical Analyst for Advanced Placement, was made available to the project (at no cost) by the Educational Testing Service. Dr. Cowell was doubly useful to the project, which he served as Chief Statistician. His extensive experience with statistics in testing enabled the project directors to make wise choices on some difficult statistical issues, most particularly the problem of combining essay and objective test scores and the problem of establishing fair cutting scores. In addition, the fact that Dr. Cowell was not only a participant but

also an outside observer helped to give the entire project additional credibility. It should be pointed out that Dr. Cowell always functioned as an advisor to the project directors and to the State University Test Officers with whom the responsibilities for making decisions always rested.

4. Essay Question Leaders. As the essay test began to be developed, it became clear that question leaders for each of the two essay questions would be essential, if the questions were to be created and graded according to the best available procedures. After considerable consultation, both within and outside the State of California, the following two appointments were made: 1) Dr. Rex Burbank, Professor of English, California State University, San Jose; 2) Dr. Gerhard Friedrich, State University Dean, Academic and Resource Planning. Each of these professors of English has had over a decade of experience participating in controlled essay readings run by the Advanced Placement Program. The professional competence and experience of the two question leaders were critically important to the success of the test program. Their responsibilities were: 1) participate in the development of the two essay questions; 2) evaluate the results of the pre-test of these essay questions and report to the project directors with suggestions for revision; 3) assist in the selection of table leaders and essay readers; 4) develop statements

for graders describing the criteria for grading; 5) select 20 to 30 sample student essays for discussion by the graders during the course of the reading and direct these discussions; 6) maintain supervisory control over the five table leaders and 30 to 35 essay readers in each question room during the reading of the essays; 7) consult with the project directors during the reading on the schedule and overall administration of the reading; 8) participate in the decisions on cutting scores; 9) prepare a final report on their responsibilities during the reading, particularly with regard to the kinds of distinctions that were made among the essays written by the students.

5. Aides. Finally, as the date of the essay reading in June approached, it became necessary to appoint a staff to insure that the test papers and other supplies were moved according to plan from reader to reader. Accordingly, a staff of 15 to 20 student assistants was selected for our use by Dr. James Ware, Chairman, Department of English, California State Polytechnic University, Pomona, and two chief aides were selected to supervise the student aides in their duties in the two question rooms and elsewhere during the reading. The chief aides were Ms. Linda Snyder, English Department secretary, California State College, San Bernardino, and Ms. Mary Reynolds, an English department secretary, California State University, Northridge. Ms. Reynolds also doubled as project secretary before, during and after the reading, a second and fulltime

job in itself. (It would have been preferable to have the project secretary remain in charge of secretarial duties during the reading, and have a separate person acting as chief aide in charge of the student aides during the reading.)

At the inception of the project, the project directors were insufficiently aware of the extraordinary complexity of a large testing program, particularly one that included the creation, administration and grading of an essay test. Fortunately, however, as the project developed and new areas of responsibility emerged, it was possible to find extraordinarly well qualified people who were willing to assume responsibility for these areas and accomplish the necessary tasks in these areas with efficiency and responsibility. Without their assistance the project could not have succeeded. Certainly, anyone undertaking to direct such a program in the future should be well aware of the specialized help he will need.

B. Meetings and Decisions

Beginning in October 1972, a series of meetings and conferences, sometimes as often as twice weekly, began among the project directors and a group of consultants. Out of these meetings a series of decisions emerged which refined and developed the ideas contained in the English Council report of October 1972 (See Appendix). The personnel and location of these meetings changed according to the agenda, but the following consultants

made themselves available for meetings at which their special expertise would be useful:

From the Chancellor's Office: Dr. David Provost, State University Dean, New Program Development and Evaluation; Dr. John Smart, Deputy Dean, New Program Development and Evaluation; Dr. David Leveille, Associate Dean, New Program Development and Evaluation; Dr. Robert Bess, Director, Academic Projects; Dr. Gerhard Friedrich, State University Dean, Academic and Resource Planning; Mr. Charles Davis, Public Affairs Associate, Public Affairs.

From the Educational Testing Service: Dr. Albert Serling, Program Director, College-Level Examination Program; Dr. Alan Seder, Program Coordinator, Berkeley; Dr. Richard Harsh, Director of the Los Angeles office; Dr. William Cowell.

The project directors were always present, representing the English departments; as the essay test began to be developed Drs. Friedrich and Burbank, along with other experienced English department representatives from the various campuses, attended several meetings. In addition, the project director was invited to spend several days at the Educational Testing Service office in Princeton, New Jersey to consult with the various offices and personnel who have been planning and administering controlled essay readings for the Advanced Placement program for many years.

While some of the decisions that emerged from these advisory meetings may have been appropriate only for this

particular testing program, it is likely that the issues and the way they were resolved will be useful as a reference for similar programs in various other states and under various other conditions. Thus, a summary of problems and decisions follows:

1. Single Closed Administration. Since we were to give an essay test and to publicize the test very widely, it was necessary for security of the test to administer the test only once. Thus, we could not use the "open" College-Level Examination Program centers in the state for our purposes, since they offer tests monthly; we needed to establish one closed administration of the test. We thus decided to offer the test on each of the nineteen campuses of the California State University and College system on May 12. This decision worked out very well. We were able to maintain complete control over all processes of the test administration and insure that test materials, including over 4,000 essay examinations, arrived at the proper place at the proper time to be prepared for the grading session. There were a few complaints from students who could not take the test the one day it was offered, but the many advantages of a single closed administration of this sort of test became apparent very quickly. Our conclusion was that under similar circumstances a single closed administration is by far the best way to proceed.

2. Location. Since the California State University and College system was sponsoring this particular test, we

decided that the test centers should be at each of the nineteen campuses in the system rather than at various CLEP centers around the state. This decision worked very well also. The test was clearly and publicly a California State University and Colleges test, not an ETS operation, and the test program redounded to the credit of the California State University and Colleges system. In addition, we needed to deal only with California State University and Colleges test officers, instead of the variety of personnel who are directing CLEP centers in various locations around the state. A few difficulties did arise: at the Pomona campus, there was a conflict with a large campus activity, and only with some difficulty were facilities found for the test. On a few other campuses there was an uncomfortable moment or two, as enrollments for the test began to accumulate and facilities were limited. We did make plans to re-direct students, in case some campus facilities were over-taxed, but in no case did it become necessary. The test officers on each campus were cooperative and wholly competent to arrange for the machinery of test taking at each of their locations.

3. Day of Test. We had originally thought that a Thursday might be more appropriate than a Saturday, since some students may have religious objections to being tested on Saturday, or be otherwise unavailable on a weekend. These reasons were counterbalanced by a possible shortage of seats available for testing on a Thursday, when classes are

in session, and by the need for high school students to obtain special permission to take a test at college on a school day. We decided therefore to offer the test on a Saturday, and that decision appears to have been wise. We heard no complaints whatsoever from students on religious grounds, and the test officers found themselves much more able to deal with problems of administration on a Saturday.

4. Date of Test. After careful consultation with representatives from the Educational Testing Service, the test date selected was Saturday, May 12, 1973. It was necessary to check carefully to avoid conflicts with other examinations the same students might wish to take. We were particularly careful to avoid conflict with Advanced Placement, since we suspected there would be students desiring to take both of these tests.

5. Time of Day. After considerable discussion, it was decided to offer the examination from 10:00-12:00 noon and from 1:30-3:30 p.m. There was some opinion that the test should be given in one block of time, say from 9:00-12:30 p.m. Such an arrangement may well have been more convenient for some students and proctors, but we decided on the later starting time in order to help students who would have to travel long distances; and we decided to break for lunch between the objective and essay tests, so that students would be refreshed and thus ready to write better essays. (The test officers were instructed to pay proctors on the basis of a single test session, even though

it was divided by a lunch break, which resolved one difficult problem in payment of proctors.) On the whole this decision proved to be successful, but there were problems. Some students showed up for one session or the other, instead of for both, though that may in part have been attributable to an unfortunate typographical error which appeared at one point in the registration form. A future administration should make clear that attendance at both parts of the examination is required in order for the examination to be considered. At least one campus took advantage of the lunch break to show some hospitality to the students on campus, offering them free soft drinks to go with their bag lunches. An enterprising campus, seeking to recruit unusually able students, could well offer the students additional hospitality. It is, of course, impossible to say how much the break between sections influenced the overall quality of the essay test, which was very high.

6. Proctors. Each of the nineteen test offices designated a person to act as chief proctor on that campus and the chief proctor appointed sufficient assistant proctors to administer the examination in accordance with normal ETS procedures. The coordination and direction of these personnel were carried out efficiently through the California State University Long Beach test office.

7. Payment of Proctors. Honoraria were paid to the California State University and Colleges institutions by CLEP according to its regular honoraria schedule. These

fees were part of the $15 test fee paid by each student and proved adequate to support the expenses of administering the entire examination. ETS would have been willing to send a single check for proctoring to the directors of the program, who would then pay the various test offices. But this idea was rejected as unnecessarily complicated. Each test office submitted to ETS its roster of students taking the test, and ETS sent checks to the individual test offices on each campus. Though there was some small delay in receiving payment, the test offices found this arrangement satisfactory.

8. Movement of Test Materials. This very complicated matter was handled with great efficiency through the ETS office in Berkeley, under the direction of Alan Seder of that office. All test materials were received in the Berkeley office in ample time so that they could be sent from Berkeley to the test centers 10 days before the administration of the test. By that time each test office knew the number of registrants and had reported that information to the test office at California State University, Long Beach; also by that time the essay questions had been printed and sent to Berkeley from Northridge, and the special test manual had been printed and delivered. Mr. Seder included an overage of 15 percent to accomodate late registrants, and all materials were received by the test centers in ample time. At the conclusion of the test, the essay examinations were mailed to Northridge by each test center; the objec-

tive tests, answer sheets, and all other materials were returned to Princeton. There were no mix-ups in the delivery of material, and the only difficulty was the late mailing to Princeton from several of the test offices, which failed to get the materials in the mail, in some cases, until Wednesday or Thursday of the week after the test. Nonetheless, all answer sheets were handled by the Princeton ETS computer according to schedule, and the objective score report was in the hands of the administrators of the program in time for the essay reading.

9. Closing Date and Late Registration. The announced closing date, by which time all registration forms were to be in the hands of test officers, was April 27, two full weeks before the test date. The test officers did, however, accept registrations up to within a few days of the test at their discretion. The established closing date worked very well to provide us with the necessary information about the amount of test materials to have available at each test center. There were few enough late registrants so that they caused no particular problem, and we received no complaints from test officers about this procedure.

10. Test Manual. The original plan to adapt a CLEP test manual, with a small supplement of our own, became impractical as we discovered the many differences between our test and usual CLEP procedures. It was thus decided to prepare a test manual specifically for our test administration. With the advice and cooperation of Alan Seder, in Berkeley, and

Ms. Betsy Barlow, in Princeton, who bore particular responsibility for the preparation of the test manual, this manual was prepared, approved, and printed in time to be delivered to the test centers with the test materials. Dr. Richard Cantey, of the California State University, Long Beach, test office, took particular responsibility for overseeing production of this manual. The preparation of a test manual should be considered as one of the special responsibilities of the coordinator of test administration; and it would be well to prepare this manual well in advance, so that the condition of urgency which prevailed during the last few days of production in 1973 need not occur.

11. Test Form. Dr. Albert Serling, Program Director for CLEP, gave instructions that the particular form of the CLEP Subject Examination, Analysis and Interpretation of Literature, which would be used in the May 12 test program, not be offered in the state of California through the regular CLEP test centers. This was a wise precautionary move to insure the security of the test program.

12. Registration Form. We could not use the standard CLEP Registration-Information form, since it asked for information of no use to us, and failed to provide the necessary information for students taking the examination. The production of a new registration form absorbed a considerable portion of the energies of the test administrators, and became the major vehicle of information about the test throughout the state. Since the form needed to be printed in great

haste in order for the publicity program to get under way, it was not possible to use the Office of the State Printer to prepare the form, and one unfortunate typographical error did appear in the final printed copy. Nonetheless, an excellent printing job was done by the Franklin Press in San Bernardino, and almost 60,000 copies of the form were distributed throughout the state. A repeat of the testing program needs to insure that the registration form be prepared well in advance, thoroughly and carefully.

C. The Objective Test

The 1972 report on equivalency testing presented the arguments for the use of the particular objective test that was used in the test administration May 12, 1973, and there is no need to review those arguments here. The California State University and Colleges English Council agreed in October 1972 that the CLEP Subject Examination, Analysis and Interpretation of Literature, was the most appropriate objective test available at that time for freshman English equivalency. This test is basically a good reading test: it contains approximately 100 multiple choice questions, all of them based on passages supplied in the test. The passages are selected so that no previous experience with them or knowledge of their background is required to answer the questions, and the passages are taken from 19th and 20th century American literature and from each of the major periods of English literature from the Renaissance to the present. The results of the test show that this objective test was

appropriate for the test group, and served effectively to make distinctions between those with college-level skills in reading literature and those without such skills. At the same time the correlation coefficient between the objective and the essay test (.477) shows that a substantial portion of what was tested in the objective test was not tested by the essay portion. Further evidence of the validity of the objective test appears when we notice that each essay question correlated more highly with the objective test than with the other portion of the essay test.

There can be no question about the use of objective testing in large scale English equivalency tests; an objective test is essential. This particular objective test worked very well and may be worth using again for that very reason. The major objection to the use of this test relates to the nature of the freshman English course itself. In many cases, essay writing in freshman English uses literature as its subject. In the May, 1973, test, since half of the total examination, the 90-minute objective test, was a test on the reading of literature, it was felt necessary that essays be written on non-literary topics. If a good objective writing test that would not be based upon an ability to read literature were to be used, it would then be possible to give an essay test on literature; such an essay test would strike many English professors as still more valid than the one given in 1973.

A new objective CLEP Subject Examination in freshman

English is now available. If this new test is used for the 1974 examination, it would make it possible for the essay test, in 1974, to be about literature. The statistics available from ETS on this new test indicate that it accurately reflects achievement in freshman English courses in a wide variety of institutions, including several in the California State University and Colleges system. But the administrators of the 1974 test will need to decide whether a new objective test, which allows them to use an essay test on literature, is worth using, when the available statistics from 1973 show that the CLEP Subject Examination, Analysis and Interpretation of Literature, accurately measures important aspects of freshman English skills.

Meanwhile the Advanced Placement Program has shown considerable interest in the California State University and Colleges English Equivalency Test, and the College Entrance Examination Board is considering possible revisions in the structure of its several college equivalency programs.

Certainly, the administrators of the 1974 examination need to consider all available instruments before a decision is made about which objective test to use.

D. The Essay Test

1. Format. It was decided that the essay examination would be 90-minutes long and be weighted equally with the 90 minutes of objective testing. There was considerable discussion of weighting the results differently, but there were as many arguments for weighting the essay more heavily than the objective test, as there were for weighting the objective

test more heavily than the essay. It was finally the consensus of all those involved that the tests should be of the same length and should receive equal weight. Various ways of distributing the 90 minutes of essay testing time were discussed, and at first it was proposed that three 30-minute essays be used, each essay to receive one reading. After further consideration and discussion, it was decided to give the students longer time to reply to two different questions, and, thus, to require two 45-minute questions. If each of the two questions received two independent readings, it seemed apparent that the reliability of reading would be increased and that a more fair score would be obtained. Thus, the test was made up of two 45-minute questions, each testing different kinds of skills. It was also decided that the time limitation would be enforced so that graders would be evaluating comparable effort; after 45 minutes, question one would be collected and question two distributed.

2. Creation of the Essay Test. Early in the planning the decision was made to create our own essay test, and not use the literary topics provided with the CLEP Subject Examination, Analysis and Interpretation of Literature. Since 90 minutes of objective testing were on literature, and since most freshman English courses in fact require writing on many subjects besides literature, it seemed fair to provide non-literary topics for the essay test. Once the decision was

made to create our own essay test, it was necessary to decide precisely what kinds of skills we were seeking to test, so that we could examine those skills. This task was accomplished by a committee that met for a full day with the project directors and the question leaders. (See Chapter 1, page 9 for the members of that committee.) After prolonged discussion, the committee decided to test on the first question the student's ability to describe an object from personal experience and his capacity to move from description to abstraction. The second question would ask for a response to two short passages, and demand the ability to respond incisively to others' ideas. The assumption was that essays emerging from personal experience call for quite different kinds of writing skills than essays comparing and contrasting quotations; both kinds of questions are common in freshman English courses. (This assumption was born out by the results; the correlation of question 1 to question 2 was only .3681. This relatively low correlation, despite the very high reliability coefficient of readers for the same question, supports the sophistication of the essay reading, which clearly did not give a high priority to elementary matters such as spelling, which may be presumed to be constant on both essays.)

3. Grading System. After considerable discussion of the various grading scales that have been used in the past, it was decided that grading would proceed on a 6-point

scale, ranging from 1 (lowest) to 6 (highest). The 6-point scale was chosen because it liberated readers from conventional A, B, C, D, and F grading pattern, and, hence, would make them more likely to abide by group standards, and because the 6-point scale contains no middle score. The graders are forced to decide between a paper that is in the top half (4,5,6) and, thus, a passing one, or in the bottom half (1,2,3) and, thus, failing. The 6-point scale is really a pass/fail scale, and the basic decision is between a middle pass (5) and a middle fail (2); since an equivalency test renders a pass/fail decision, the 6-point scale was more appropriate than any scale that contained a middle score. At the same time, simpler versions of the pass/fail (a 2- or 4-point scale) seemed to allow insufficient scope for student variation. The results of the use of the 6-point scale were very satisfactory, and we recommend that such a scale be used for such tests in the future. When the four grades were added, the lowest possible score for a paper that responded to both questions was a four, and the highest was a 24. The results of this scale are set out in considerable detail below. All that need be said here is that the 6-point scale worked very well for the purposes of this examination and allowed us to make the distinctions that were needed with confidence.

4. Question 1. The following question was distributed to all students:

"Many observers of our society claim that modern man, immersed in materialism, is 'owned by his objects.' Yet many of us have objects that we treasure not just for their material value but for a variety of other reasons. Assignment: Describe one or more objects which are important to you. Explain what values they represent and comment on those values."

The following directions for scoring were distributed to all readers engaged in the grading of question 1:

"The student should be rewarded for what he does well in response to the question. Here the student is asked to describe one or more objects important to him. He is further asked to explain what values they represent and to comment on those values. He is told to think about the question and to plan his response.

"Note that the question does not demand that the object or objects be unusual ones.

"Essays which misinterpret 'objects' as 'objectives,' and which deal mainly with generalized abstractions (life, God), should be read sympathetically, but they should ordinarily not receive above 2, since they fail to understand and properly respond to the question.

"Possible Scores:

6 A superior response will not just name one or more object but describe them in some detail, and it will not just identify the values represented but explain and comment on them, their nature and their source. A superior paper will be literate and orderly.

5-4 These scores will be useful for a well-handled paper which is deficient in one or two characteristics of the superior response, i.e. in description of the object or objects and in explanation of the values represented, but which is otherwise competently written.

3-2 These scores will be useful for the following kinds of papers:

--those in which only one part of the two-part question is addressed;

--those in which the representativeness of specific objects is ignored;

--those which treat the subject in superficial or stereotyped fashion;

--those in which the writing exhibits several weaknesses, in wording and other respects.

1 This score is to be used for papers which are lacking in focus and substance, depart from the assigned topic, and/or exhibit serious writing faults.

* Non-response papers and papers which do not fall into the foregoing categories, extensively argue with the question, or are otherwise idiosyncratic, should immediately be brought to the attention of the table reader and the question leader."

The following report was prepared by Dean Gerhard Friedrich, question leader for the first question:

SUMMARY REPORT ON QUESTION #1

For the first, experimental administration of a system-wide equivalency test in Freshman English, it was decided to employ a ninety-minute composition portion consisting of two distinct writing tasks. The first essay question would be relatively open-ended, permitting the candidate to start writing without any "hang-ups"; the second essay question would be more structured, requiring the candidate to deal with a given subject.

A planning committee called by the co-directors considered a variety of test questions. Two were eventually selected for careful rewriting, pretesting, and further refinement. A special effort was made to assure that the phrasing of the questions would be absolutely clear, with the practices of the Advanced Placement English examination as a guide. Pretest responses from students in comparable California

State University and Colleges classes were reviewed to determine the actual suitability of the proposed essay questions and particularly the wording of the instructions. The entire composition portion was then administered to a larger sample of California State University and Colleges students, hopefully for norming purposes.

After the candidates' essay booklets had been received from the test centers, the two question leaders read a large number of responses from all the test centers and selected sample essays illustrating the range of performances to be identified, from excellent to incompetent. It was decided to employ a 6-point scale, with scores of 6, 5 and 4 indicating degrees of creditable performance, and scores of 3, 2 and 1 degrees of deficiencies in terms of college-level composition. Both question leaders participated, immediately before the reading of the English Equivalency Test at Kellogg-West, in the Advanced Placement English reading at Rider College, and spent considerable time in drawing up "rubrics," i.e. guidelines or criteria for scoring essay responses, in accordance with Advanced Placement models. Copies of the final version of Question #1 and of the instructions to readers of Question #1 responses are attached for reference.

All readers of the candidates' compositions were drawn from among the English faculties of the nineteen campuses in The California State University and Colleges. Readers were divided into groups of six or seven, each with an experienced table leader. The table leaders for each question were brought together in a pre-reading session to harmonize grading standards on the basis of a representative sample of papers previously selected. Subsequently, together with their question leader, they similarly instructed their respective readers. Consistency of standards was further ensured by having the table leaders regularly double-check scores assigned by readers; readers were also encouraged to confer with their table leaders on any scoring problems. Question leaders in turn double-checked the scores assigned by table leaders, and from time to time throughout the reading polled the entire group on additional sample papers. In this way, a workable consensus was rather easily achieved and maintained. In the relatively few instances in which the scores assigned to a Question #1 essay were at least two points apart, the essay received a third reading, followed by discussion and appropriate adjustments.

It should be noted that the reading of Question #1 essays was both fair-mindedly responsible and very efficient. Readers took the task of assigning appropriate scores seriously and managed to read more essays per day than had been anticipated. The candidates' compositional abilities

covered an amazingly wide range, from brilliant to illiterate, and in most instances the readers of the same composition assigned identical scores. Moreover, in many instances a candidate performed at the same level in response to both essay questions. However, there were also striking exceptions, when no correlation was apparent between a candidate's responses to the two essay questions, nor between the essay portion and the performance on the objective test.

Attached are examples of candidates' responses to Question #1, illustrating the variety of objects chosen as well as the scored levels of performance.

Gerhard Friedrich
June 30, 1973

The following student responses to Question #1 were sample papers used during the reading to illustrate the grades on the 6-point scale:

SCORE OF ONE

"I am going to write on some object which are pretty important to me. The first would be my religon because I think I should put my faith in something other than 'man', and worldly goods. I also believe that I as a person have that right to look forward to something bigger and better in life, I an live life better day by day. This is important because I'm a person who doesn't like to have things cramed down my throat then expected to digest it easily. I feel I have the right to choose who and what I believe in, without someone handing me 2 alternatives to chose from. I want to choose freely on my own will and judgement. I feel that I am old enough to choose and have 'free choice'. Another object or value which is important to me is my 'Freedom'. I like to do what I like, go where I want, see what I want, to a point where I do not interfer with another persons 'Freedoms.' My whole life is based on what I can do for me and other people. Without this freedom I could not serve to the best of my ability my fellow man. I couldn't put forth 100% because I would be restricted to do only certain things. Therefore, without my religon and freedom to do what I want I should have never have been born and with out these two basics of life I do not see how any man can live. So I have told you two of my basic values, again they are 'Freedom to choose my own religon' and basic Freedom to live an everyday life."

SCORE OF TWO

"Heat, exhaust, fumes, burning rubber and smoke are all caused by a remarkable invention that has spured our society into being one of the most materialistic in this modern age. Our society today depends on the car for transportation. We overlook the bad side of this invention for all the wonderful things the car has done for us.

No longer are people confined to one small region for their entire lives. Trips to the coast or to a distant city for a day are not unheard of now. It has actually broadened our horizons for we can meet new people, go new places. People we haven't seen in along time are in easy reach.

Working days are shortened with the use of the car. Instead of walking many miles to work, it provides fast and easy transportation on highways.

Status is related to owning a car. Some people seem to feel that the bigger a car is, the better it is. Socio-economic status is based on the number of high value materialistic things we own. Having four or five cars in a family tends to raise a family's status.

Cars come in all shapes and sizes. Big or small we can find one that fits the needs of everybody. Compact, economy and luxary cars are priced to fit people with even low income budgets as well as high.

There is a limit that people using cars must draw. Excessive use of a car can damage our enviornment. Taking a car into high mountain area can damage or even ruin flora and fauna. Pollution from cars cannot be stopped unless all cars are banned.

With all the good and bad sides to cars, which way can we turn? Cars can be used for destructive purposes as well as useful, meaningful reasons."

SCORE OF THREE

I

"Bear-Bear"

"Blue body with a white tummy and round black eyes, soon to be loved. This poor little teddy bear went through so much just for me.

When I came home after my birth, I had 2 sisters and a strange new friend waiting for me. My sister Monnie Leigh had a teddy bear placed in my crib. They tell me I actually giggled when I first saw my teddy bear.

Well days went on and after countless washings, due to being thrown-up on or thrown into the toilet, my little teddy bear had seen his last day with those round black eyes. Yes,

my little teddy bear needed some new eyes and blue buttons seemed perfect. After all, blue eyes match a blue body!

Monnie Leigh decided, one day, that she would teach me how to say Teddy Bear. It was a noble effort on her part even though all I could repeat was Bear! The Teddy bear still didn't have a name.

The next day when I was in the living room, I realized that my teddy bear wasn't around. With the terrifying thought that he might be gone, my tiny voice piped up with; 'Bear-Bear, Bear-Bear!'

There was no mistaking what I wanted. Now that he had a name, Bear-Bear would never be more than a helping hand away.

I never really pondered on why I kept Bear-Bear all these years. It's simple - I love him and I'll keep him many more years I imagine. Bear-Bear will always mean love, security, and friendship. Perhaps that's why I choose my friends carefully and value their love so greatly. My friends have been wonderfully good to me (and vise-versa) since the very first time my giggle said 'Hi, let's be friends!' "

SCORE OF FOUR

"At this point in life, I don't have many objects which I value a great deal. The few things I do value have a sentimental value rather than a monatery value.

The first objects I value are trophies I won diving. They have a very deep sentimental value to me because it was my reward for the hours and years I practiced, working for a goal. When I look at my trophies on the shelf I think of all the joys I felt at winning and also the heart break of losing. I think of the self-control and self-discipline I gained at going to practice each day while my friends were at the beach. This, right now, is very important to me. These trophies have no real monetery value or sentimental value to anyone but me because only I earned these trophies and only I cherish them.

Another object which I value is a ring I recieved from my grandmother. This ring is also a sentimental object to me. I recieved it after she passed away and so it is my rememberance of her. This ring also has monetery value. It is a gold ring with a small diamond in it. Others would value it because of it's worth but my family and I are the only ones who value it for a sentimental reason.

My next valuable object is my wallet. My value on my wallet is very sentimental. In it, it contains all my pictures of friends and experiences which I have gone through. If someone stole my wallet or I lost it, I would rather lose my money than some of the pictures inside.

Maybe this is because I am not overflowing with money at the moment but right now my pictures come over my money.

My last valuable is a watch I received at Christmas from my boyfriend. This is very important to me because it is from him. I have had watches before from my parents and they really didn't have very much sentimental value to them. But my watch contains many memories and I would really be upset if I misplaced it. Just like the ring, it is valuable to others because of the cost but to me it is the person who it signifies.

These are my most valuable objects and it isn't because of the amount of money they're worth but instead for the sentimental value of them. I don't really own anything excessively expensive so I really don't value many things for they're value in money."

SCORE OF FIVE

"As I look back on my life, the object that I place the most value on is the house that I grew up in. For sixteen years I walked through its doors and lived in its rooms. That house became a part of me.

Now, almost nineteen years old, it stands in a middle-class suburb of Los Angeles. The surrounding streets are lined with well-kept homes and neatly-trimmed yeards.

Children that I don't know play baseball on the avenue, and cars that I don't recognize fill the driveways. My dear house is in an alien world both to me and to it.

My family took pride in that home. We bought it new, put in all the landscaping, and made it a beautiful place to live. In all the years we lived there, I never once took its loveliness for granted. I would sit and look at it and know what a wonderful home we had.

That long avenue was my world. Little playmates moved in and out of the other houses, but I was the stable one. I didn't believe we would ever leave our home.

That building saw my first step, heard my first word, and watched me fall off of my first bicycle. It stood by when I was sick and was there for all the happy moments too. It became more like a person, part of the family.

Leaving it all alone for new people to run about in was next to impossible. Is it as lonely as I am?

Our new house is bigger and more modern than that one was. Still, this makes no difference. It will never be home."

SCORE OF SIX

"We have in our living room a music box, which for three generations has given pleasure to the eyes and ears of my family. It stands about a foot high and measures about two feet in length and width. Except for a spray of flowers carved on its face, the outside is unadorned. Inside, pasted to the lid, is a turn-of-the-century lithograph of a pair of plump cherubs. There is a set of tin records, perforated here and there, that goes with the music box. It is run by winding it up and releasing the spring. Its tunes are dated; 'My Gal Is A High-Born Lady' and 'I Guess I'll Telegraph My Baby', haven't been among the top ten for quite a while, but this does nothing to lessen the enjoyment they give.

My grandfather was the first to own the music box. He traded a horse for it and presented the music box to my grandmother as a gift. They had been married for only six months. At first it was a very big deal. A music box in a Nebraskan farming town can cause quite a commotion, but as time went by and the popularity of 'victrolas' grew, the music box passed into oblivion.

Ignored and dusty was the way my father discovered it in the cellar. He cleaned it up and got it running and showed it off patronizingly to his friends as a relic from his parent's youth. It was played at parties as a novelty, but again it lost out against the incoming rage: the radio. So back into the cellar went the music box to await rediscovery one more time.

This time it was my sister and I who resurrected it. We hauled it out into the light, dusted off its rosewood sides and listened to the songs first heard what seemed to us to be eons ago.

My grandparents grew old, and being practical people, decided to divide their possessions with their children before their death to avoid a tragic scrabble afterwards. To my father went the music box, and he carefully brought it to our home and revived it one more time.

Now, despite its years, it keeps on playing its old familiar songs. I love the old music box. It can never be associated with a price tag. My grandfather acquired it with an honest trade and it has been handed down through the years. The music box symbolizes my grandfather's love for my grandmother, my father's years at home, my sister and I exploring in the dark cellar and countless fine memories. I love it for its beauty, the rich, soft red of the rosewood, the way it gleams in the sun. I love the whirr of the motor and vigorously cranking the handle. I love the corny song titles and running my fingers over the rough surface of the records. And although its value is largely sentimental, its worth stems from the fact that is has survived many years with grace and beauty; something very few <u>people</u> can claim."

5. Question 2. The following question was distributed to all students:

"A Founding Father said: 'Get what you can, and what you get hold;
'Tis the Stone that will turn
all your Lead into Gold.'
A contemporary writer said: 'If it feels good, do it.'

Assignment: What do these two statements say? Explain how they are alike and how they are different."

The following directions for scoring were distributed to all readers engaged in the grading of Question 2:

"6-POINT SCALE

Key to Scoring English Equivalency Essay Examination

The student is asked to write an essay in which he explains what the two statements mean, and how they are alike and how they differ. He should be rewarded for what he does well in his response to the assignment. Papers should be scored for their overall quality.

An extremely well-written response may be scored a point higher than it would be scored on the basis of content alone. A poorly written response may be scored a point lower.

Spelling errors should not ordinarily be counted against the score.

POSSIBLE SCORES:

6 A superior response will be a well-organized essay that does the three things asked for in the assignment. It will explain the meanings of both quotations and compare and con-

trast them; it may explain the meanings by means of comparison and contrast, or it may explain the meanings and compare and contrast them. The best essays will note that while the quotations share a concern for self rather than others, the first emphasizes results or consequences and acquisition of achievement, while the second emphasizes gratification without regard for consequences. The best essays may well mention the figurative meanings in the first quotation or see an objective-subjective distinction between them. An essay getting a six score will show a high degree of competence generally, though it may have minor imperfections. It will support generalizations with appropriate details or examples.

5-4 These scores will apply to responses that concentrate more on one quotation than on the other, or that deal with both subjects somewhat less thoroughly than the essays scoring 6. Essays in this group may have minor errors in writing.

3-2 Papers in this category deal with both quotations but may:

--be lacking in supporting details or examples, or treat both quotations superficially;
--give adequate attention to one but too little to the other;
--fail to see similarities in meaning between the two and to make distinctions between them;
--misunderstand or misinterpret the meaning of either or both;
--be primarily critical or argumentative rather than expository;
--have serious faults in writing;
--drift away from the topics or reveal considerable irrelevancy.

1 This score should be given to any response that is on the topic but has no redeeming qualities.

Non-response papers and papers that are completely off the topic should be given to the table leader."

The following report was prepared by Dr. Rex Burbank, question leader for Question #2:

SUMMARY REPORT ON QUESTION #2

The two essay questions were devised by subcommittees at our first meeting in April. Question 2 was designed to suggest an organizing principle upon which the examinee could build a 45-minute essay. Two quotations were used, and students were asked, first, to explain their meanings and, second, to tell how the meanings were alike and how they differed. The structure suggested was thus based upon explanation or definition, comparison, and contrast.

Having decided upon the question, the subcommittee arranged with Dr. Lid to have it pretested by a group of 25 freshman English students at California State University, Northridge, under his direction. The subcommittee developed a scoring key that set forth agreed-upon standards for grading the pretest samples.

When the pretest papers came to me several days later, I read and scored them in accordance with the key and compared the essay scores with the grades given those freshman students on the first two essays assigned in their course. There was enough of a correlation between scores and grades to suggest that the question would generate essays by high school seniors that could be scored meaningfully on a 6-point scale.

We (Dean Friedrich, Dr. Lid, Dr. White and I) met on June 5 in Northridge, and on that day Dean Friedrich and I read in excess of 100 test papers, written by students who had taken the test in May, and selected twenty-five each (he for Question 1, I for Question 2) for samples. The samples were marked and Xeroxed for use at the reading in Pomona from June 16-20. During the following week Dean Friedrich and I both attended the Advanced Placement readings at Rider College, Princeton, New Jersey, he as an observer and I as a participant. In the evenings we went over our chosen samples again, analyzed them, and wrote our scoring keys to reflect the qualities found in the test papers. The scoring keys proved, with minor exceptions, to be both accurate and useful to the readers. A copy of the scoring key for Question 2 is appended to this report.

On Saturday morning, June 16, Dean Friedrich and I met separately with the table leaders assigned to us. Each of us was to supervise the reading of responses by five tables of readers, who were to convene in the afternoon. I passed out 12 sample responses to my five table leaders, and we went through each sample, scoring it and discussing the scores in relation to the test papers in order to arrive at agreement as

to what qualities or weaknesses would be found in responses at all points on the 6-point scale. By noon, the table leaders were in close agreement on their grading--their scoring, done without prior discussion, was within a point of mine in most cases.

In the afternoon, the readers met with table leaders and repeated the scoring of samples as we had done it in the morning. Though both the table leaders and the readers were inexperienced in this type of reading, the 'training session' went well enough for us to begin the actual reading of 'live' test booklets at 4 p.m. We continued the next morning but not until after four more samples had been read did we resume reading 'live' ones. Always, the readers were told, the goal was uniformity in scoring: every student taking the test, we emphasized, had a right to feel that his test was being graded in the same way and by the same standards as everyone else's. Readers were asked, again, to sacrifice their own grading policies and standards for those established by the scoring key and by the group as a whole. Most readers understood the need to do this and cooperated.

Samples were passed out periodically for the next two days. Readers scored them and their scores were compared with those agreed upon by the table leaders and me. The readers were asked to adjust their scoring in accordance with the samples. In addition to sampling, checking was done by having table leaders gather papers at random from those already read and scored by readers at their table, scoring the papers themselves, and recording both their own scores and those of the readers on a 'check sheet.' The table leaders (who read and scored papers without seeing the scores given by the readers) brought the checklist with the test booklets to me, and I then read and scored them myself without looking at the scores given by the table leaders and their respective readers. Thus I was able continually to check the scoring of the readers with that of the table leaders and both against my own. Where trends developed at some tables -- such as the tendency to grade too high or too low or to settle in the 3-4 range -- they could be and were corrected by passing out samples representing the full range of scores in order to remind the readers again to use the full scale and get clearly in mind once more the qualities of papers at each point in the scale. A high degree of agreement was achieved, and in the vast majority of cases the two readings of Question 2 were within one point of each other. When there was a spread of two points a response was read a third time by an individual in a special group of our best readers chosen for this task. Papers (and there were remarkable few) with a spread of 3 points were given <u>two</u> readings by that group. Most papers with a spread of 3 points were radically uneven in quality and so the discrepancies were understandable. I'm satisfied, however, that generally the papers were graded with a very high degree of uniformity, reliability, and validity.

We didn't expect perfection from even the best of the student responses, but for a score of 6 we required that the paper show a high degree of writing skill in doing what was asked by the assignment. The essay was expected to say something, directly or indirectly, about the meanings of the quotations and compare and contrast them. The best papers were expected to demonstrate the ability to build on the implied organization with meaningful details, to develop the general points the writer made, and show a high degree of understanding of the quotations. The best papers, for instance, revealed a consciousness that the first quotation looked to the future, to consequences or results, while the second emphasized immediate gratification of desires or wishes; the first was figurative and stressed acquisition, the second was literal and suggested personal satisfaction. Students writing the best papers accurately perceived one or more such distinctions.

A paper given a score of 5 fell just short of the 6 essay in having minor faults in writing, being slightly less well developed, or displaying a slightly less acute understanding of one of the quotations; or emphasizing one quotation over the other. A paper given a 4 differed only in degree from those given a 5; it was awarded a 4 rather than a 3 in that, overall, it suggested competence, despite whatever minor faults it might have, rather than incompetence.

We asked the readers to decide first of all, as they read, whether the paper was an upper or a lower half (that is, a 6-5-4 or a 3-2-1)essay, to look at the quality of the paper as a whole, first, and then to make the necessary distinctions within those two categories. We asked them to forget the letter-grading they are used to in their own teaching and to remind themselves that it was essential to use <u>both</u> ends of the scale as well as the middle; only then could we make relative judgements and pertinent distinctions.

Papers given a 3 grade were lacking in details, or gave <u>too little</u> attention to one of the quotations, or failed to perceive similarities in the quotations, were primarily critical rather than explanatory, lacked unity, or had serious faults in writing. A 2 paper had one or more of these weaknesses in greater degree than one given a 3. A paper was to be given a 1 if it was on the topic but was so badly written that it suggested illiteracy or clear incompetence. We gave a 0 to off-topic essays or papers with no response at all. Examples of papers given scores 1 to 6 are appended to this report, as are the forms used in the readings."

The following student responses to Question #2 were sample papers used during the reading to illustrate the grades on the

6-point scale:

SCORE OF ONE

"Get what you can, and what you get hold; tis the stone that will turn all your lead into Gold.

This statement attributes to the young, while working hard in life one may establish a foundation and from this point molitier this foundation into a future which in later years will prosper with prosperity and fortune."

SCORE OF TWO

"These two statements are from two widely separated times the first was spoken in a time when life was short and hard when most peoples were poor and the wilderness was sometimes just behound your door.

The other was writen at a time when just the reverces was true life is long and for the most part easy when most are prospress and to find adventure one must go looking for it rathir then having it come chasening after you.

As they come from two diffrent ages they represent two diffrent views of happyness the first is a life time of happiness not only for you but for your kids. The other is a day to day kind it seems to lack the resposably of the other, it does not say anything about the resultes on others the first statement does not say this things eather but the implication is there. It depends very much on you point of view. Thou in maney ways that diffure in the end thay both mean the same to be someone you must have something only in this way will all the troubles of you life have meaning and be worth it and all your lead be turned into Gold."

SCORE OF THREE

"The practical aspects of human motivations should not be confused with immorality. The Founding Father offered sound advise. Paraphrased it means: One may attain his realistic goals by diligently working at one's maximum level. The contemporary author says: One should wear his morals like a suit of clothes; whatever comfortable is fine, only they should not tight under the armpits.

No basis of comparison, I believe, can find common philosophical points between the two statements. I, therefore, find the question a loaded one. Only if one were to misunderstand the

true intentions of the Founding Father could there be a likeness comparison, fore the contemporary author's statement is easily understood, and not apt to be confused.

Perhaps analysis of the statement is indicative.

'Get what you can, and what you get hold.

'Tis the Stone that will turn all your lead into Gold.'

Today, frequently, English speaking people confuse the meanings of words and generally speak poorly. The word 'can' in the first line is a modal auxiliary, and when used as such means 'ability'. If the line were to contain 'may,' then the advise would not be refering to a goal that could be acheived, but rather to 'get what you may' -- which connotes immorality of dishonesty. 'Get what you can' therefore means, 'get what you have the ability to get. Although the 'can' could connote immorality it probably does not, because the erudite Founding Father undoubtedly knew the rhetorical difference between 'can' and 'may' and therefore did not make a mistake in syntax.

The second line:

'Tis the Stone that will turn all your Lead into Gold.'

refers to the Philosopher's Stone of Alchemy. This seemingly would role out the idea that the Founding Father's message is concerning morality, or anything neccessarily intangible. It seems rather obvious that the what one can get and hold must be material possessions. In this aspect, then, the statement by the contemporary author has little relation at all to the Founding Father's statement!

The statements have little relation to each other. To say that they do would mean that anything practical is immoral."

SCORE OF FOUR

"The first quote says to find all opportunities, and when you find them, keep them because they will bring fortune to you.

The second quote says to do the things that make you feel "good" or happy. Things that make you feel worthwhile are what you want to do.

The two statements are similar because they both want the reader to get ahead and find happiness. The first wants you to find success, and through success, richness (joy or happiness). The second wants you to find joy and happiness in doing something that comes natural to you.

The two statements have differences, not only in their words, but in their ways of obtaining happiness. The first tells the reader 'Get what you can' but doesn't put restrictions on how to get it. In the second verse of the first statement 'will turn all your Lead into Gold' could be taken by the reader to mean that all opportunities will lead to good while the author, in general reading, seems only to be saying always strive for better and don't let an opportunity pass you because of indifference or indecision. The second statement could also have

a bad connotation for some readers, such as critics, because it could be taken to mean do something even if it is wrong, simply because it makes you happy.

In general reading, though, the first statement would seem to be an enlonged version of saying the same thing as, the second statement."

SCORE OF FIVE

"The two statements, one by the Founding Father and one by a contemporary writer, are very similar, and also very different.

The basic thought in both of these is to get or do what you want. If you have a goal, or an object that you want to obtain, push and shove, or squirm and wriggle, but work towards your goal to get it. This is one of man's basic impulses -- the writers are just encouraging the fulfillment of man's wishes.

The thought in the statement made by the Founding Father is the foremost example of the idea of getting an object. It seems that the writer urges man to acquire materialistic things such as land, wealth, or possessions. He urges you to get it; and once you get it, to keep it.

The contemporary writer seemed to be speaking more of spiritual or nonmaterialistic things. He seems to urge that man express himself freely in his speech, his actions, his life style. Love can be more easily meant in this statement than in the one by the Founding Father. This contemporary statement is used quite often in regards to the 'New Morality.' No one really knows what <u>that</u> is, but everyone wants to be a part of it, and the phrase, 'If it feels good, do it', has become a sort of cliche.

The Founding Father and the contemporary writer both seemed to want to express the idea that man must look after his own interests, be they pleasures or businesses.

Whether or not the two writers had any of the same ideas in making their statements is not known, but the similarities indicate that it is very possible. I'm sure the thoughts of the Founding Father were more <u>conservative</u> than those of the contemporary writer.

These two statements both say a lot about man's way of life, now and in the past. The acquisition of materialistic and spiritualistic pleasures has always been a principle part of our life."

SCORE OF SIX

"People today are looking at life in an entirely different perspective than those who witnessed the birth of this nation.

Life is no longer a constant struggle for survival (as it was then), but instead it is a time for people to expand and grow in knowledge and understanding. Today, people live their lives by an entirely different philosophy than our ancestors of yesteryear.

A Founding Father once said, 'Get what you can, and what you get hold; 'Tis the Stone that will turn all your Lead into Gold.' The philosophy he is expounding is that of being a 'go-getter,' a person who can get all he can. But it is also a philosophy of values. This means that one should also value what he achieves or gets, and never let go of it. In those days things of value were often hard to come by, so they valued things greatly. Because it might be that one little item that could 'turn all your Lead to Gold,' or make you prosperous. It is a liberal philosophy in that it says to go out and get whatever you can, yet it is conservative in that it says to hang on to it and not to let go. It is a philosophy of hard work, and being appreciative of that work.

But the people of today's society have great difficulty following a philosophy of this nature. We aren't living in 'hard times', at least the majority of Americans aren't, so it is hard to follow a philosophy that old, or seemingly dated.

A contemporary writer states, 'If it feels good, do it,' and it appears as though people are 'doing it', because we are living in a completely different way that our Founding Fathers did. People today don't feel the need to hang on to things in hopes of it producing great prosperity. Life is too 'easy going' and the struggle for survival is non-existant. So people decide what they want to do, and they do it. There are usually no great risks involved (as there were earlier) and in general, life has turned into more of a game or routine for people than a struggle or a challenge.

The two philosophies do have some common ground. They both say to go out and do, act, achieve, get, accomplish. They both call for action, but one says to hang on to the dividends and the latter of the two says to move on to something else that 'feels good.'

I feel the latter of the two philosophies shows the real moral deterioration of our society today. Life has become so routine, so mechanical, so dehumanized that it seems to be almost lifeless. And so a contemporary writer tells people to do what feels good. What feels good is not always what is best, and often times we have to do things that don't feel good, but they must get done. If one always does what feels good, a lot of problems go unsolved and new ones are created.

I favor the first philosophy because it puts value in what we do and it puts purpose in life: that purpose being working hard and appriciating the end result. It may not always feel good, but will provide self-satisfaction in knowing that something had to get done and you did it."

6. The Norm Sample. One of the central problems in the test program was the establishment of norm scores. The objective test had been normed in 1966, on a group of students from 17 schools. It was not altogether clear that that norm sample was in fact representative of the State University population, although recent experience with the CLEP Subject Examination, Analysis and Interpretation of Literature, at the University of Texas has validated the norm scores. Ideally, a new norm sample taken from California State University and Colleges students completing a year of freshman English would have supplied new norms for both parts of the examination. It proved impossible to find the time and money to provide such a norm sample for the objective test, and we decided to stand with the norm population achieved by the CLEP program in 1966 for the objective test. However, we did seek to obtain a norm sample from California State University and Colleges students for the essay test.

The results of our attempt to acquire this norm sample were disappointing and not really reliable: we had been promised almost 600 test papers from 12 different institutions; we in fact obtained 259 papers from 7 institutions, and almost half of the papers represented the State Polytechnic Universities of Pomona and San Luis Obispo. It also became clear that the motivation of the students writing the norm sample at the end of their class work was far below that of the students taking the test for credit.

The norm sample did show, however, a correlation between scores achieved on the essay and projected student grade in the course, and, hence, tended to validate the accuracy of the essay grading. If the norm sample were to be considered valid, it could be used to show that the students given credit through the testing program generally write better than the students receiving credit through freshman composition classes. But that would be a risky conclusion to draw in the light of the project's failure to accumulate a sufficiently representative and well motivated norm sample on the essay portion of the test. (See Table I and Table II, pages 56 and 57.)

Future attempts to gather a norm sample for this purpose will need to take into account the problems we optimistically felt we could solve. We made personal contact on each of the 12 campuses with directors of two-term freshman English courses, and each department chairman had agreed to participate in the norming project. We accompanied the request for participation with the letter which follows and a packing sheet (see pages 69 and 70) in order to assist the instructors in giving both the anticipated course grade and a forced ranking of his students (in case he had theoretical difficulties with the conventional grading system). It was our expectation that the personal contact, the general willingness of English departments throughout the system to cooperate with the project and the very specific instructions contained with the norming materials would all lead to a

TABLE I

PARTICIPATION IN THE NORM SAMPLE

Colleges Agreeing to Participate	Sections	Booklets Returned for Use in Norm Sample
Dominguez Hills	2	47
Fresno	2	12
Hayward	1	--
Humboldt	1	--
Long Beach	3	49
Northridge	3	--
Pomona	2	49
Sacramento	1	--
San Francisco	2	--
San Jose	1	22
San Luis Obispo	4	67
Stanislaus	1	20
TOTAL	23	259

669 Test Booklets were distributed; 259 were returned; 242 were useable for purposes of norming. The useable return rate is approximately 36 per-cent.

TABLE II

ESSAY TEST GRADES ACHIEVED BY THE NORM SAMPLE,
ARRANGED BY PROJECTED COURSE GRADE

Projected Course Grade	Essay Mean	Standard Deviation	Total Number
A	13.5	4.0	26
B	12.9	3.0	75
C	10.6	3.5	98
D	10.1	3.2	33
F	7.8	2.4	10
ABC	11.9	3.6	
DF	9.5	3.2	
		TOTAL	242

CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

Northridge, California 91324

DEPARTMENT OF ENGLISH

May 14, 1973

Dear CSUC English Professor:

We appreciate your willingness to participate in the norming of the English Equivalency Test, which was administered to over 4,400 prospective entering students on May 12, 1973. The norm sample that you will help collect is absolutely crucial to the proper grading of the essay tests, and we are asking you to read and follow the instructions on this sheet with great care.

The only essays we can use are essays written by students completing two terms of freshman English. Please tell the students taking this examination that the grades on the essays will be part of their course grades and please do all you can to encourage them to do their best writing. But if you intend to make any marks on the papers or to return them to the students, please make a Xerox for such use. We must receive the essay booklets unmarked in order to be able to use them. Allow the students exactly 45 minutes for each essay. Be sure not to open the essay questions until you administer them. All answers must appear in the enclosed green CLEP Essay Answer Books, and no student should need more than one book for both essays (thus you need to redistribute the green books if you use two separate class hours). If necessary, a second book may be used.

The cover sheet of the answer book must be indistinguishable from the high school books from the May 12 test, if we are to use your papers; thus, ask your students to use their home address (not a dormitory), to fill in for Subject "English Test," and to insert your college or university name for Center Name. If a student does not know his social security number, he may leave that item blank. Above all, have the students enter May 12, 1973 for date, no matter what date you may in fact give the test.

We are also asking you to give us some essential additional information about each student. Please fill in the enclosed packing sheet carefully, for without that information the papers you send us cannot be used. The last column "Comparative Ranking" needs some explanation. Choose the three or four weakest students in the class, and write "1" for their ranking; choose the three or four best students and call them "6." The average students should be split between 3 and 4; use no minus or plus signs. This will allow you to provide a ranking from 1-6 of all your students, without regard to grades.

It seems clear from the nationwide reaction to our test program that the success (or lack of success) of what we do will have a profound impact on national testing programs and on the future of essay testing in such programs. As far as we know, this is the first time English professors have assumed responsibility for the reliability and total administration of such a test, and the program thus becomes an example of professional activity without precedent. The scrupulousness and fairness of our procedures will be subject to careful scrutiny, and your part in these procedures is central to the entire grading process.

Return the entire packet not later than June 5 to Professor Richard Lid, Department of English, California State University, Northridge, Northridge, California 91324.

Sincerely,

Edward White

Edward White, Director
California State University English
Equivalency Test

EMW/mr
Enclosures

CSUC ENGLISH EQUIVALENCY TEST

Packing Sheet for Norm Sample

Faculty Member _____ Course Number _____
College or University _____ Number of Tests _____
_____

Student Name (Alphabetized)	Social Security Number	Estimated or Final Course Grade (not an examination grade)	Comparative Ranking (see cover letter) (use 6-point scale)

reliable, useable, and heavy return. As Table I shows, these expectations were not well founded. The rather poor participation, even by English departments who had agreed to participate in a test they themselves were sponsoring, is very clear evidence of the residual resentment most English departments continue to feel towards equivalency testing.

University of California Norm Sample. Two sections of students completing two terms of freshman English at the University of California wrote the essay test, and their papers were graded during the essay reading session. As with the California State University and Colleges norm sample, the graders could not distinguish these test papers from the others; the University of California sample was coded and later separated.

Twenty-four students from University of California Los Angeles received a mean score of 14.29 (standard deviation 3.54); 16 students from University of California Riverside received a mean score of 13.56 (standard deviation 3.31). The total University of California sample of 40 students achieved a mean score of 14.0 (standard deviation 3.43), coincidentally the passing score set for the test group.

7. The Essay Reading. The experience of the Advanced Placement program based on substantial ETS research on essay grading, has shown that for an essay reading to be reliable, it must take place under controlled conditions. Thus, an early decision was made to organize an essay reading

session of approximately one week, during which essay graders would be gathered together and work together in order to achieve group grading standards which would be fair to the students who had written the essay test. After a survey of local accomodations, it was decided to make reservations at the Kellogg-West Conference center, at California State Polytechnic University, Pomona for the week of June 16-20, 1973. That conference center supplied accomodations for about 75 staying at the center, working space (two large rooms for the ten tables of readers; and three smaller rooms for supplies, statistical work and the posting of grades), and excellent food within the state-allowed per diem for college professors. The date was chosen in order to allow ETS sufficient time to score the objective test and return completed information on the scores to the test administrators for use during the essay grading session. (ETS also supplied two complete sets of mailing labels, so that notification of the results could be mailed to the students speedily.)

The professional competence of those reading the essays is crucially important to the reliability of an essay reading. Since the project was sponsored by the California State University and Colleges English Council, the project directors requested each of the nineteen English Department chairmen to nominate from the full-time English Department staff at his campus those who would be both willing and able to participate in group essay reading at

the time scheduled. Since the English Department chairmen nominated the readers, and screened them for their ability to participate, it was a relatively simple task to appoint readers. The care with which these nominations were made is reflected in the reliability statistics of the reading, among the highest ever achieved. Seventy-five readers were appointed, and, of these, the ten who were the most highly recommended, or who had had previous successful experience with controlled essay reading, were appointed as table leaders responsible for insuring the reliability of reading at one table. The question leaders (See pages 17-18) coordinated and supervised the work of the five table leaders in each room, and directed the discussions of grading criteria.

It was difficult to estimate how many readers should be appointed. Not until the end of April was it known how many examinations would be graded, and there was no way of knowing the speed at which the reading would progress. We estimated that a single reader would be able to read and score ten papers an hour, and that an eight hour work day would allow for seven full hours of reading with time for rest and discussion of standards. We, thus, estimated a single reader could read and score 70 papers a day, or 280 during the four days of the reading. As it turned out, we underestimated the capabilities of our highly professional readers, who, despite their lack of experience, in fact were able to read and score 120 papers per day with great

accuracy.

In order to insure fair reading of the papers, several steps were taken in preparation: a) Randomizing, to insure that no single batch of papers would come from the same location. Essay booklets from the nineteen test centers were randomly arranged into batches of 25. Each batch of 25 was given a batch number and batch cover sheet to insure that it moved properly through the reading. b) Labeling. Special labels were given to us by the ETS office in Berkeley. The labels had four detachable squares which would allow scores to be recorded through carbon on the test booklet, and yet be concealed from subsequent readers. At the conclusion of the fourth independent reading, the carbon paper backing was torn off and all four scores were exposed to be tabulated. c) Coding. All papers from the norm sample were coded on the back cover so that they could be separated from the test papers and given separate statistical treatment. The norm papers were then inserted into odd numbered batches from 1 through 55 so that they would be easily separable from the others.

At the start of the reading, one reader aide (a local, commuting, student assistant) was assigned to each table. One was assigned to each of the chief aides, and one assigned to the statistician. After the first day, it became clear that additional student aides were needed, and through the office of the English Department chairman at Cal-Poly Pomona, additional names were provided to us. Thus, in addition to the original thirteen aides, at the peak of the reading an

additional six aides were being used, along with a few readers who were assigned to work with the readers' aides in the posting of grades.

When the four readings of the essays had occurred, and the scores were revealed, there were occasional booklets on which large differences (2 points or more) appeared in the grading of one, or both, of the essays. These papers were referred back to the question leader, who either reconciled the two grades himself, or assigned the paper to an experienced reader for reconciliation. An additional group of papers, from students with particularly high objective test scores, but low essay scores, was also referred to the question leaders for reconsideration, and a few of them were rescored as passing.

The management of the essay reading is a huge undertaking, requiring a substantial amount of time and attention if it is to go well. A staff of approximately 100 individuals was employed to conduct the reading at the conference center, and on the whole the reading went efficiently and successfully. The project directors did underestimate the amount of labor and coordination involved in running the essay reading properly. But, in part due to the efforts and cooperation of the management of the conference center, the reading was concluded successfully.

E. Converting the Essay Grade and Combining the Scores.

The objective test is scored on the familiar ETS scale which converts the raw score to a scale ranging from 20

(low) to 80 (high) with a mean of 50 and a standard deviation of 10. Since we decided to weight the essay portion of the test equally with the objective portion, it became necessary to convert the essay scale (which ranged from 4-low to 24-high) to the same scale as the CLEP scale so that the scores could be combined. After a prolonged period of consultation with statistical experts, it was decided to accomplish this score conversion by the equal percentile method. The following statement by William Cowell summarizes the procedure used:

> "'Equating' scores means making them equivalent with respect to some characteristic. Here, the essay scores were transformed to the same scale (20 to 80) as the objective scores so that the transformed essay scores could be added to the objective scores to obtain a composite score in which the two parts would be equally weighted. Had the raw essay scores been added to the objective scores, the essay part would have had a weight of only 1/4 of the total. Also, the transformation was done in such a way that the converted essay scores can be directly compared with objective scores. That is, a converted essay score of 48, for example, represents the same level of performance as an objective score of 48.
>
> "Of several acceptable methods of equating which are available[1], the equipercentile (curvilinear) method was selected as the most appropriate. The converted score corresponding to each raw essay score is that score in the distribution of objective (CLEP) scores such that the percent of the group scoring below that objective score is equal to the percent of the group scoring below the given essay score. For example, to find the converted score corresponding to a raw essay score of 15, we observe from the distribution of essay scores that about 68 percent of

---

1. Angoff, W.H.; 'Scales, Norms, and Equivalent Scores' in R. L. Thorndike (Ed.) Educational Measurement, 2nd Ed., Washington, D.C.: American Council on Education, 1971, 563-590.

the group scored below 15 on the essay part. From the distribution of objective scores it is determined that a score of 53 is that score below which 68 percent of the group scored. So a raw score of 15 is equivalent to a scaled score of 53. Similarly, a scaled score is found for each raw essay score. Table III shows the converted score corresponding to each raw essay score.

"V. Technical Notes about the Equating.

(1) The equating was done using data from the Equating sample ... The sample was used instead of the total group so that the statistical work could proceed along with the essay reading. Score conversions and distributions could be prepared for the sample data in time for a meeting with Question Leaders at the reading. The time and cost of equating was also reduced by using a sample rather than the total group. Scrambling the essay booklets before the reading provided a sample which was representative of the total group.

(2) A method of graphical smooting, described in 'Educational Measurement'[2], was used to interpolate between percentile ranks in the distribution of objective scores.

(3) Although the equipercentile method was selected because it allows for a non-linear relationship between objective and essay scores, the correlation between raw essay and converted essay scores is .999. This extremely high coefficient of linear relationship[3] indicates that the simpler linear methods of equating could have been used with essentially the same results as the equipercentile method.

---

2. Ibid., page 572.

3. Ibid., page 513.

See the Appendix for a series of statistical reports prepared by Dr. Cowell on the project, including a full discussion of the problem of equating essay and objective test scores.

TABLE III

SCORE CONVERSION TABLE

CSUC ENGLISH EQUIVALENCY TEST

MAY 1973

Essay Total Score	Converted Score
24	76
23	73
22	71
21	69
20	67
19	64
18	62
17	59
16	56
15	53
14	51
13	48
12	46
11	43
10	41
9	38
8	35
7	33
6	31
5	28
4	26
3	24
2	22
1	20
0	18

F. Determination of the Passing Score.

One of the most difficult decisions was to establish the passing score on the essay test. One possible passing score could have been a total of 16, that is, a paper which was given a score of 4 on the 6-point scale by each of the 4 readers. However, the 16 converted to a score of 56 on the

20 to 80 scale and clearly represented a level of performance above that which freshman English professors normally expect of their students at the end of the course. On the other hand, it would have been possible to set the passing score at 11, since the "C" students in the norm sample (see p. 68) were obtaining essay ratings of about that level. Once again, the conversion of the 11 to the CLEP scale (43) demonstrated that that score was clearly too low and that the grade of "C" reported for these norm sample papers was not an accurate reflection of their ability.

After much consultation and discussion, the following committee agreed unanimously that we should set the passing level at a score of 14, and that a minimum passing level of 13 should also be set: Professors White, Lid, Burbank, Friedrich; Test Officers Abbott, Cantey, Bradley (from Cal-Poly, Pomona); Statistician William Cowell.

The arguments for a passing score of 14 were as follows:

1) that score represented an average grade of two 4's and two 3's; that is, two of the college professors grading the essay felt that the essay to be passing at the lowest level, and two felt it to be failing at the very highest level of failure.

2) the score of 14 converted to 51 on the 20 through 80 scale which was very close to the mean "C" score on the objective test of 49.

3) the norm sample, in so far as it could be relied upon,

showed that the mean score for students receiving A, B, or C was 11.9, and, when awarding college credit, it is well to err a bit on the side of severity.

4) the mean score of the University of California students completing a year's work in freshman English, a sample apparently better motivated and better administered than our own norm sample, turned out to be precisely 14 on a group of 40 papers.

The same arguments were used to justify a minimum passing score of 13, which converted to 48. A student could not pass the test with an essay score below 13, and if he or she had a score of 13, an objective score of 52 or above (which was achieved by 32 percent of the test group) was necessary for a total passing score.

G. Notification of Results. Each student who took the test received a score report from ETS giving the results of the objective test. An insert was included in that notification, pointing out that an additional report, which would give the results of the entire examination (including the essay test) would follow.

The administrators of the test program prepared two letters, which were sent to all students. The intention of the letters was to stress the meaning of the examination, to test college-level ability, and, particularly, to stress that those who did not pass should not think of themselves as "failures."

The following letter was sent to those who did pass:

> "Dear _____
> On behalf of the English faculties in the California State University and Colleges, we congratulate you for successfully passing the Freshman English Equivalency Test you took May 12, 1973.
> You are among a select number of students who have demonstrated the same level of ability many students achieve only after a full year's work in college English. In particular, you have shown that you have college level competence in writing essays and reading literature, a competence we urge you to develop by taking the more advanced English classes you are now entitled to enter.
> The test you passed consisted of the ninety-minute subject examination, Analysis and Interpretation of Literature (part of the College-Level Examination Program of the Educational Testing Service), and a ninety-minute essay test composed and graded by State University English professors. In order to pass, you needed to perform satisfactorily on both parts, and achieve a combined score equal or superior to that of the average "C" student who has completed a year's work in Freshman English.
> You should take this letter with you when you register for classes at any of the nineteen California State University or College campuses; it will allow you to

claim six semester units of college credit for what you know and permit you to enroll for courses with Freshman English as prerequisite. You may want to consult with the English department chairman about which English courses will be most appropriate for you.

If you do not plan to enroll in the California State University or Colleges, you should still save this letter, for it is the only notification of your success that we can send you. Institutions outside the State University system may well give full or partial credit for your passing score, and you should inquire about your chances of gaining credit.

Please be sure to complete the enclosed information card and return it immediately so that our records can be complete."

The following letter was sent to those who did not pass:

"Dear _____

Even though you were not among the group of students who passed the Freshman English Equivalency Test, May 12, 1973, you should not consider that you "failed" the test. Your level of achievement on the test did not match that of the average college student who has completed a year's work of college English; those who passed the test were able to accomplish on their own what everyone else does in a year of college class work. Only those who combine natural ability at English with careful and systematic college level study have a good chance at such college equivalency credit, and it is no surprise that many able high school students did not pass.

Remember, the test was not an aptitude test, measuring your native abilities at English; it was an equivalency test, measuring your level of achievement. Thus you may be very talented at reading and writing--you probably are, since you elected to take this test--but the test results showed that you have not yet developed these talents at the college level.

Your total test score will not be sent to anyone or any institution, even on request, and the only record that you took the essay test will be kept for research purposes at the English Test Center. (Your score on the objective portion of the test may be sent by the Educational Testing Service, only on your request, wherever you wish.)

When you return the enclosed information card, you will make it possible for us to do this research; we

want to know more than we do about how to help able and ambitious students prepare themselves in English, and we need to evaluate the results of the 1973 test in order to prepare better ones in the future. We appreciate your help in this research, and we urge you to continue to follow your own interest in writing and literature as you go through college."

During the summer, the office of Institutional Research, in the Chancellor's Office, prepared a roster of those who passed. This official roster was sent to each California State University and Colleges admissions office, and was used to corroborate student applications for credit on individual campuses. In September, after the University of California had decided to award credit on the same basis as the California State University and Colleges system, one additional roster was sent to Berkeley to be used by the University of California system.

A few requests for information about the examination have come from private institutions, whose students have passed the examination and sought credit. These requests have been dealt with on an individual basis. In addition, a number of secondary schools have requested information about their students' scores. If the program expands, the handling of test results will need to be reconsidered, since these problems will surely grow.

H. Public Information.

"What if they gave a war" the bumper sticker says, "and nobody came?" What if we gave a test and nobody came? Throughout the development of the test program we were haunted by the fear that the program, so expensive, so painstakingly developed, so carefully prepared, would not attract enough students to make it worth the giving. Certainly high school students, particularly the most able ones, are subjected to an extensive series of tests in preparation for college: they take the various college board tests, Advanced Placement tests, various aptitude tests, and, of course, their various final examinations in their classes. How were we to draw the attention of these able high school students to yet another test program, and motivate them to spend $15 to take this new kind of test?

It thus became clear by the middle of January, 1972, that a large public information effort would be necessary, if the students who stood to profit from the English Equivalency Examination were to enroll for it. We planned a two-month campaign in March and April to attract the attention of students to the new opportunity for college credit we were offering them. This campaign encompassed the following activities:

1. Each high school counselor listed on the Chancellor's office list received 5 registration forms and one poster in the mail.

2. Each member of the California Association of Teachers of English received one registration form and one poster in the mail.

3. The Counselor's Digest issues of February 1973 and April 1973, featured major articles on the front page about the English Equivalency Examination. Every high school counselor received these issues.

4. Each of the nineteen California State University and Colleges Admissions Offices was requested to include one registration form for the test along with information to admitted students during the month of March and April.

5. The Office of Public Relations, in the Chancellor's Office prepared several press releases which were distributed to newspapers, radio and television stations around the state. These releases led to a major article in the Los Angeles Times, and widespread newspaper, radio and television coverage of the announcement of the test in early March. The director of the program was interviewed on KNX New Radio in Los Angeles, and television announcements were made in the San Francisco Bay Area and in San Diego. Local campus news releases followed, in many cases, and a series of announcements appeared in various media during March and April.

6. A few of the California State University and Colleges English department chairmen took special pains to notify high schools in their vicinity of the new

opportunity for gaining college credit, and these chairmen sent registration forms to high school English department chairmen with personal notes attached.

This public information campaign was costly and time-consuming, but absolutely essential to the success of the project. The number of registrants for the examination, over 4,000, was beyond our expectations and ample evidence that the program met a genuine need; many of those who could profit from the program found out about it. Nonetheless, during the month of April and in subsequent months, we kept hearing of students or entire high school classes who knew nothing about the test program and were disappointed at not being able to participate in it. The public information campaign did not reach all who should have heard about the examination, and publicity efforts should be improved upon in the future. The weakest point in the dissemination of information about the test seems to have been the failure of high school counselors to communicate information to high school English teachers, who, in turn, would encourage their best students to enroll for the examination. Perhaps articles in high school student newspapers would solve this particular problem.

Nonetheless, an extraordinary amount of publicity was generated for the test program in a very short amount of time. If the program continues, it should become generally known, and a regular part of the spring testing cycle for high school seniors.

Then efforts will need to be made to attract adults, very few of whom took the examination.

IV

FOLLOW-UP STUDIES

The Office of Institutional Research in the Chancellor's Office has put all information about the students tested on computer tape, and has assumed responsibility for conducting follow-up studies in order to determine the effects of the examination upon the careers of those students who took it and either passed or failed. Along with the notification letter, every student who took the examination received a return post-card requesting certain information about his training and his plans. Approximately 30 percent of these post cards were returned and the results of that survey are reproduced on pages 75 - 76.

That office will continue to collect information about the students who were tested, and particularly about those students who will be entering and studying in the California State University and Colleges during the next several years. The results of this longitudinal study will be published later. Among the questions to be answered by these follow-up studies are: 1) will students use the 6 units gained by the examination to abbreviate their college career, or rather to advance the level of their studies? 2) will students receiving credit for freshman English by examination avoid further English courses or rather be encouraged to take more

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES
English Equivalency Examinations - Spring 1973
Descriptive Statistics - Post Card Response

Total Response = 1,234 = 30.3% of Total Applicants

Response From Those Passing = 1,216 = 98.5% of Total Response

Response From Those Failing = 18 = 1.5% of Total Response

Response From Those Passing = 1,216 = 89.3% of Total Passing

Fall 1973 Educational Plans - Those Passing

California State University and Colleges	=	88.6%
California Community Colleges	=	4.5%
University of California	=	3.5%
Other Colleges or Universities	=	1.4%
California Public High School	=	.2%
No Response - This Item	=	1.8%

Institution Last Attended - Those Passing

Public High School	=	78.5%
Private High School	=	13.7%
California Community Colleges	=	3.5%
California State University and Colleges	=	2.7%
University of California	=	.2%
Other Colleges or Universities	=	.3%
No Response - This Item	=	1.1%

Taking "Enriched" English Course in High School
Those Passing

YES = 35.3%

NO = 63.5%

No Response - This Item 1.2%

CSUC - Institutional Research
November 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES
English Equivalency Examinations - Spring 1973
Distribution of Fall 1973 Applications - Those Passing

Campus/segment	Number	Percent
Bakersfield	19	1.6
Chico	75	6.2
Dominguez Hills	7	.6
Fresno	30	2.5
Fullerton	113	9.3
Hayward	46	3.8
Humboldt	102	8.4
Long Beach	71	5.8
Los Angeles	18	1.5
Northridge	91	7.5
Pomona	59	4.9
Sacramento	41	3.4
San Bernardino	9	.7
San Diego	120	9.9
San Francisco	80	6.6
San Jose	94	7.7
San Luis Obispo	106	8.7
Sonoma	48	3.9
Stanislaus	10	.8
All Campuses	1,139	93.7
California Community Colleges	59	4.9
University of California	98	8.1
Other	34	2.8
Total Applications	1,330	109.4

CSUC - Institutional Research
November 1973

advanced English courses? 3) will students who receive credit for freshman English by examination achieve as well in advanced courses requiring writing as students with approximately equal ability who took the freshman English course? 4) Do students with the particular skills and abilities to pass this kind of test tend to go to or come from certain schools or certain programs? 5) to what degree is success on this sort of examination dependent upon race, economic level, or other extra-academic factors?

V

THE 1974 ENGLISH EQUIVALENCY EXAMINATION

As this report goes to press, word has been received that an examination similar to the 1973 examination will be supported by the Chancellor's Office, and the same administrative staff will direct the CSUC examination for the English Council on May 11, 1974.

Administration of a large-scale testing program is an enormously complex, enormously time-consuming, enormously expensive, yet enormously rewarding project. It is our hope that the program will in fact become a normal part of the procedures of higher education, and that college bound students with the ability to profit from college-level studies will find opportunities in high school to develop their skills in English as far as they are able. It is our hope and expectation that such students will be tested by English department faculty who are well aware of what is expected in their freshman courses, will be certified by

these faculty as having met these demands, will receive credit for what they have learned, and will move into more advanced courses with the enthusiasm and delight that successful students bring to a subject they love. Such a result can benefit the teaching of English at both the secondary and college levels and can only be rewarding to students and faculty alike.

VI

CHARTS AND TABLES

A. Distribution of Composite Scores - Total Applicant Population

B. Distribution of Composite Scores - Applicants who Passed

C. Distribution of Composite Scores - Applicants who did not Pass

D. Distribution of Converted Essay Scores - Total Applicant Population

E. Distribution of Converted Essay Scores - Applicants who Passed

F. Distribution of Converted Essay Scores - Applicants who did not Pass

G. Distribution of Objective Scores - Total Applicant Population

H. Distribution of Objective Scores - Applicants who Passed

I. Distribution of Objective Scores - Applicants who did not Pass

J. Intercorrelations Among Part and Composite Scores - Total Applicant Population

K. Intercorrelations Among Grouped Essay and Composite Scores - Total Applicant Population

L. Male Applicants by Age Category

M. Female Applicants by Age Category

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

English Equivalency Examination - Spring 1973

Distribution of Composite Scores

Total Applicant Population

Score	Frequency	Relative Percent	Cumulative Percent
51-59	12	0.3	0.3
60-69	96	2.4	2.7
70-79	452	11.2	13.8
80-89	908	22.4	36.2
90-99	1064	26.3	62.5
100-109	836	20.6	83.2
110-119	445	11.0	94.1
120-129	170	4.2	98.3
130-139	52	1.3	99.6
140-149	13	0.3	99.9
150-152	2	0.1	100.0
Total	4050*	100.0	

Mean = 95.363 Median = 94.485 Mode = 91.000

Standard Deviation = 14.869

*21 Applicants did not complete Essay Examination

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES
English Equivalency Examination - Spring 1973

Distribution of Composite Scores for those Applicants who Passed

Score	Frequency	Relative Percent	Cumulative Percent
100-104	357	26.2	26.2
105-109	331	24.3	50.5
110-114	237	17.4	67.9
115-119	201	14.8	82.7
120-124	99	7.3	89.9
125-129	70	5.1	95.1
130-134	37	2.7	97.8
135-139	15	1.1	98.9
140-144	11	0.8	99.7
145-149	2	0.1	99.9
150-152	2	0.1	100.0
Total	1362	100.0	

Mean = 111.442 Median = 109.407 Mode = 101.000

Standard Deviation = 9.547

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

English Equivalency Examination - Spring 1973

Distribution of Composite Scores of those Applicants who did not Pass

Score	Frequency	Relative Percent	Cumulative Percent
51-54	4	0.1	0.1
55-59	8	0.3	0.4
60-64	21	0.8	1.2
65-69	75	2.8	4.0
70-74	166	6.2	10.2
75-79	286	10.6	20.8
80-84	402	15.0	35.8
85-89	506	18.8	54.6
90-94	557	20.7	75.3
95-99	507	18.9	94.2
100-104	116	4.3	98.5
105-109	32	1.2	99.7
110-114	5	0.2	99.9
115-119	2	0.1	100.0
120-121	1	0.0	100.0
Total	2688*	100.0	

Mean = 87.222 Median = 88.202 Mode = 91.000

Standard Deviation = 9.472

*21 Applicants Did Not Complete Essay Examination

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES
English Equivalency Examination - Spring 1973

Distribution of Converted Essay Scores
Total Application Population

Score	Frequency	Relative Percent	Cumulative Percent
22-26	11	0.3	0.3
27-31	67	1.7	1.9
32-36	273	6.7	8.7
37-41	630	15.6	24.2
42-46	903	22.3	46.5
47-51	988	24.4	70.9
52-56	654	16.1	87.1
57-61	218	5.4	92.4
62-66	217	5.4	97.8
67-71	75	1.9	99.7
72-76	14	0.3	100.0
Total	4050*	100.0	

Mean = 47.916 Median = 47.662 Mode = 48.000

Standard Deviation = 8.402

*21 Applicants Did Not Complete Essay Examination

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

English Equivalency Examination - Spring 1973

Distribution of Converted Essay Scores of those Applicants who Passed

Score	Frequency	Relative Percent	Cumulative Percent
48	171	12.6	12.6
51	237	17.4	30.0
53	218	16.0	46.0
56	262	19.2	65.2
59	188	13.8	79.0
62	113	8.3	87.3
64	86	6.3	93.6
67	38	2.8	96.4
69	26	1.9	98.3
71	9	0.7	99.0
73	8	0.6	99.6
76	6	0.4	100.0
Total	1362	100.0	

Mean = 55.910 Median = 55.135 Mode = 56.000

Standard Deviation = 5.800

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

English Equivalency Examination - Spring 1973

Distribution of Converted Essay Scores of those Applicants who did not Pass

Score	Frequency	Relative Percent	Cumulative Percent
22-26	11	0.4	0.4
27-31	67	2.5	2.9
32-36	273	10.2	13.1
37-41	630	23.4	36.5
42-46	903	33.6	70.1
47-51	580	21.6	91.7
52-56	174	6.5	98.1
57-61	30	1.1	99.3
62-66	18	0.7	99.9
67-69	2	0.1	100.0
Total	2688*	100.0	

Mean = 43.868 Median = 44.130 Mode = 46.000

Standard Deviation = 6.369

*21 Applicants Did Not Complete Essay Examination

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

English Equivalency Examination - Spring 1973

Distribution of Objective Scores
Total Applicant Population

Score	Frequency	Relative Percent	Cumulative Percent
22-26	11	0.3	0.3
27-31	80	2.0	2.2
32-36	290	7.1	9.4
37-41	665	16.3	25.7
42-46	843	20.7	46.4
47-51	883	21.7	68.1
52-56	633	15.5	83.6
57-61	364	8.9	92.6
62-66	209	5.1	97.7
67-71	75	1.9	99.6
72-76	18	0.4	100.0
Total	4071	100.0	

Mean = 47.448 Median = 47.140 Mode = 43.000

Standard Deviation = 8.854

CSUC - Institutional Research

September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

English Equivalency Examination - Spring 1973

Distribution of Objective Scores of those Applicants who Passed

Score	Frequency	Relative Percent	Cumulative Percent
45-47	113	8.3	8.3
48-50	218	16.0	24.3
51-53	283	20.8	45.1
54-56	186	13.7	58.7
57-59	194	14.2	73.0
60-62	164	12.0	85.0
63-65	91	6.7	91.7
66-68	70	5.1	96.8
69-71	26	1.9	98.8
72-74	15	1.1	99.9
75-76	2	0.1	100.0
Total	1362	100.0	

Mean = 55.577 Median = 54.661 Mode = 52.000

Standard Deviation = 6.318

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

English Equivalency Examination - Spring 1973

Distribution of Objective Scores of those Applicants who did not Pass

Score	Frequency	Relative Percent	Cumulative Percent
22-26	11	0.4	0.4
27-31	80	3.0	3.4
32-36	290	10.7	14.1
37-41	665	24.5	38.6
42-46	799	29.5	68.1
47-51	544	20.1	88.2
52-56	216	8.0	96.2
57-61	74	2.7	98.9
62-66	25	0.9	99.8
67-71	4	0.2	100.0
72-73	1	0.0	100.0
Total	2709	100.0	

Mean = 43.361 Median = 42.951 Mode = 43.000

Standard Deviation = 6.914

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES
English Equivalency Examination - Spring 1973

Intercorrelations Among Part and Composite Scores
Total Applicant Population

	Essay #1		Essay #2					
	Read. 1	Read. 2	Read. 1	Read. 2	Essay Total	Essay Conv.	Objective	Composite
Essay #1								
Read. 1		.6951	.3013	.3083	.7580	.7550	.3564	.6348
Read. 2			.3156	.3176	.7617	.7573	.3663	.6439
Essay #2								
Read. 1				.6824	.7561	.7513	.3833	.6532
Read. 2					.7577	.7573	.3685	.6434
Essay Tot.						.9946	.4816	.8472
Essay Conv.							.4777	.8429
Objective								.8642

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES
English Equivalency Examination - Spring 1973

Intercorrelations Among Grouped Essay And Composite Scores
Total Applicant Population

	Question 1	Question 2	Essay Total	Essay Conv.	Objective	Composite
Question 1		.3681	.8253	.8214	.3923	.6944
Question 2			.8253	.8204	.4096	.7068
Essay Total				.9946	.4816	.8472
Essay Conv.					.4777	.8429
Objective						.8642

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

English Equivalency Examination - Spring 1973

Male Applicants by Age Category

Age Category	Failed			Passed			Total		
	N	Rel.%	Cum.%	N	Rel.%	Cum. %	N	Rel.%	Cum. %
16 and under	1	0.1	0.1	2	0.5	0.5	3	0.2	0.2
17	57	4.4	4.5	20	4.8	5.3	77	4.5	4.7
18	1051	81.5	86.0	322	78.0	83.3	1373	80.7	85.4
19	152	11.8	97.8	47	11.4	94.7	199	11.7	97.1
20	10	0.8	98.6	4	1.0	95.7	14	0.8	97.9
21	3	0.2	98.8	2	0.5	96.2	5	0.3	98.2
22	3	0.2	99.0	3	0.7	96.9	6	0.4	98.6
23	2	0.2	99.2	1	0.2	97.1	3	0.2	98.8
24	1	0.1	99.3	0	0.0	97.1	1	0.1	98.9
25 - 29	5	0.4	99.7	6	1.5	98.6	11	0.6	99.5
30 - 34	2	0.2	99.9	2	0.5	99.1	4	0.2	99.7
35 and over	2	0.2	100.1	4	1.0	100.1	6	0.4	100.1
All Ages	1289	75.7	-	413	24.3	-	1702*	100.0	-

*Age Data unobtainable from 15 (M or F) Applicants

CSUC - Institutional Research
September 1973

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

English Equivalency Examination - Spring 1973

Female Applicants by Age Category

Age Category	Failed			Passed			Total		
	N	Rel.%	Cum. %	N	Rel.%	Cum.%	N	Rel.%	Cum.%
16 and under	2	0.1	0.1	2	0.2	0.2	4	0.2	0.2
17	130	9.2	9.3	74	7.8	8.0	204	8.7	8.9
18	1136	80.6	89.9	754	79.8	87.8	1890	80.3	89.2
19	117	8.2	98.1	69	7.4	95.2	186	7.9	97.1
20	7	0.5	98.6	9	1.0	96.2	16	0.7	97.8
21	2	0.1	98.7	6	0.6	96.8	8	0.3	98.1
22	1	0.1	98.8	2	0.2	97.0	3	0.1	98.2
23	2	0.1	98.9	1	0.1	97.1	3	0.1	98.3
24	0	0.0	98.9	6	0.6	97.7	6	0.3	98.6
25 - 29	4	0.3	99.2	4	0.4	98.1	8	0.3	98.9
30 - 34	1	0.1	99.3	5	0.5	98.6	6	0.3	99.2
35 and over	8	0.6	99.9	12	1.3	99.9	20	0.8	100.0
All Ages	1419	59.9	-	944	40.1	-	2354*	100.0	-

*Age Data Unobtainable from 15 (M-F) Applicants

CSUC - Institutional Research
September 1973

APPENDIX

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES
ENGLISH EQUIVALENCY EXAMINATION

1973

EQUIVALENCY TESTING IN COLLEGE FRESHMAN ENGLISH:

A Report and a Proposal

The English Council
of
The California State University and Colleges

October 1972

TABLE OF CONTENTS

PREFACE, By Edward M. White 1

1. EQUIVALENCY TESTING: THE CENTRAL ISSUE 3

2. EQUIVALENCY TESTING IN FRESHMAN ENGLISH 5

3. STRENGTHS AND WEAKNESSES OF OBJECTIVE TESTING 7

4. STRENGTHS AND WEAKNESSES OF ESSAY TESTING 12

5. EQUIVALENCY TESTING IN FRESHMAN ENGLISH IN THE CALIFORNIA STATE UNIVERSITY AND COLLEGE SYSTEM 13

A. Aims and Objectives of Freshman English 15

B. The Test: Objective and Essay 15

C. Administration of the Test 16

1. Proposed New CLEP Freshman English Test: Fall 1974 16

2. Analysis and Interpretation of Literature: Fall 1973 19

3. Essay Grading: Supervision and Expenses 21

4. Passing Scores 22

5. The Use of Test Scores 23

6. The Colleges and The Schools 24

6. HOW EQUIVALENT IS EQUIVALENCY? 25

Appendix I: Expense of Essay Reading

Appendix II: Objective Tests in English

Appendix III: List of Correspondents

PREFACE

By

Edward M. White
Chairman, Department of English
California State College, San Bernardino

In the fall of 1971, the California State Colleges began large-scale equivalency testing for entering freshmen at two colleges, using tests developed by the College-Level Examination Program (CLEP). (This program, sponsored by the College Entrance Examination Board [CEEB], is administered by the Educational Testing Service [ETS].) After the results had been publicized, serious professional evaluation of the validity, scoring, and administration of the tests began among the faculties; the State College English Council raised a number of objections to the English Composition General Examination in particular, as well as to various aspects of the program in general. The Chancellor's Office proved receptive to the English Council's objections, and to other questions raised by a series of statewide committees and subcommittees that have considered the Fall 1971 program. In late spring of 1972, the Chancellor's Office agreed to support a summer study to be undertaken by a committee of the English Council, to investigate equivalency testing in the area of English and to recommend an appropriate program for use by the now renamed State University and Colleges.

This report is the result of that study. It is not exhaustive, since such a task in this area would have demanded far more time and support than was available. It is an attempt to focus the major issues in such a way as to point to their solution, and it recommends a method of equivalency testing in English which is responsive to our discipline and practical to implement.

This report has passed through a series of drafts and presentations which have made it, in its present form, an expression of the best thinking of the English Council as a whole--perhaps even of the English profession as a whole. Since Spring 1972, when the Council directed me to prepare this report, I have consulted widely with English department and freshman English chairmen throughout California, and have corresponded, sometimes at considerable length, with over two dozen specialists in the field elsewhere in the United States and in England. I have kept citations to a minimum throughout the report, which is written for laymen as well as for the professional, so I must thank here the many teachers, writers, and scholars whose published work and whose substantial and thoughtful letters to me have contributed to our findings. I owe a particular debt to Professor Jess Ritter of California State University, San Francisco, who worked closely with me throughout the study, and to Dr. Albert Serling, Program Director for CLEP, who spent a week in San Bernardino to give us the benefit of his wide experience. The English department chairmen and faculty who participated in the Southern California Advisory Meeting, August 3, 1972, and in the Northern California Advisory Meeting, September 14, 1972, will notice the many improvements made in the report as a result of their suggestions. I am also grateful for the advice given me by William Schaefer, Executive Secretary of the Modern Language Association; Robert Hogan, Executive Secretary of the National Council of Teachers of English; and most particularly Michael Shugrue, Executive Secretary of the Association of Departments of English, who first helped me discover where to turn to dispel my previous condition of happy ignorance about the entire area of testing in English.

1. EQUIVALENCY TESTING: The Central Issue

Equivalency testing has become common practice in higher education, and has long been widely accepted, at least in theory, in English departments. All but two of the forty-six four-year California colleges and universities responding to the 1971 Association of Departments of English Freshman English Survey, for instance, indicated that there was a way to exempt students from freshman English at their institutions. In addition, the Advanced Placement Program, also administered by ETS for CEEB, is widely accepted as equivalent to college work; a score of 3, 4, or 5 is accepted as six semester units of college credit throughout the State University and College system. (See a memo entitled "Systemwide Policy on Advanced Placement and Credit" sent by Vice Chancellor Langsdorf to all State College Presidents, June 16, 1971.)

But only recently has equivalency testing been open to very large numbers of students. Advanced Placement candidates, for instance, are relatively few in number, able and ambitious students, from a limited number of secondary schools; they enroll in specialized courses, and generally perform better than college and university students on their examinations and in their subsequent college work. Nonetheless, AP originally encountered considerable faculty resistance, and has become widely established and accepted only within the last decade. The CLEP program has greatly expanded opportunities for college credit by examination and hence has once again focused attention on the major theoretical issue raised by such credit. But since such large numbers are involved, the arguments have become particularly heated.

Those who argue for such testing assert that it benefits the individual. No one should be asked to repeat work in college that he has mastered; he should receive credit for what he knows and proceed to appropriate levels of learning.

Those who argue against such testing also assert that the needs of the individual are primary. To substitute mechanical tests of competency for the individual search for excellence is in fact to cheat the student of possibilities for individual growth.

These arguments, which can be and have been developed at great length, and which lead to rhetorical heights of passion, point to the practical weaknesses in both positions. Certainly college courses ought not to be rote repetition of what is already known, and certainly equivalency testing ought to lead to more advanced learning. When faculty argue against equivalency testing without much knowledge of available tests, or when testing people proclaim the uselessness of college course work without knowledge of the innovations and expansion of freshman studies, the conflict becomes severe. (In an article on CLEP, The College Board News, May 1972, claims the five general examinations afford freshmen "the opportunity to eliminate one entire year of study and expense," which is a strange and sad way to speak of what is available in higher education.) There is plenty of blame to go around for a quarrel which is essentially foolish, and for which students and higher education in general must suffer.

As in so many heated theoretical arguments, both sides are right, since they are talking about different things. Some of the tests that have been used are in fact poor and invalid; no one sensible defends them. Some college courses have apparently not been worth the taking; no one really defends them. But we need not and should not take extreme positions. No one could argue against a program of equivalency testing that satisfies these two conditions:

1) the tests must be in fact college level ones, valid for their stated purposes, and properly normed--in short the tests must gain academic respectability similar to that won by the Advanced Placement program, and

2) the purpose of the tests must be so clearly seen that no one can take them as a way to cheat students of their education by huddling them through credits to save cash; the tests need to be administered so that they in fact help students develop their fullest individual capacities.

Everyone stands to benefit from equivalency testing responsibly done.

2. EQUIVALENCY TESTING IN FRESHMAN ENGLISH

The issues discussed in Section One are more or less applicable to all fields of study, but they are most pronounced in the area of freshman English.

It is no wonder that equivalency testing in freshman English is a long-standing problem. The course itself is a long-standing problem, nationwide. It is the most widely required college course (in 1970, 93.2 percent of all four-year colleges and universities required at least one term of English), and a million or more students enroll in freshman English each year in this country. Yet there is relatively little agreement nationwide about what should be in such a course; while the most generally accepted intention is to improve students' ability to write, English teachers now use a large number of different approaches, none of which is demonstrably certain of success. Since the course is itself in such an unsettled state, it is no wonder that so many of the testing programs are confused in purpose and in content.

The sharpest problem for freshman English courses is one that relates directly to the issue that divides us about equivalency testing: is the objective of the course some kind of minimum competence, what Albert Kitzhaber called "immediate therapy for students whose academic future is clouded by their inability to manage the written form of English"? Or is the primary purpose "to focus the student's attention on fundamental principles of clear thinking and effective expression of that thinking" (Albert Kitzhaber, Themes, Theories, and Therapy, 1963, pp. 2, 3.). The view of English as "therapy," as fulfilling its function by imparting correct spelling and other conventional forms of expression, is widely held outside of the profession and even by 48.9 percent of the English departments in the United States (according to Thomas Wilcox, reporting on The National Survey of Undergraduate Programs in English, in College English, 6 [March 1972], 688). This is the view of freshman English assumed by most placement tests, with their heavy stress on error-hunting and supposedly correct expression. But over half the profession nationwide and all the English departments in the California State University and College system reject this vision of freshman English, in favor of Kitzhaber's second view. Correct knowledge of formal English, valuable as it is for many purposes, is not all that is taught in our classes; hence such knowledge is not alone sufficient for equivalency. Our freshman English courses are more concerned with developing an awareness of the various levels of usage, which are appropriate to various situations, than in abstract notions of correctness; and we are far more interested in helping students develop and test ideas in writing than in maintaining the supposed purity of the tongue.

Since freshman English has such varying objectives and definitions, we should not expect any single national test, however reputable, to satisfy

the profession as a whole. We, however, need to insist that tests designed to examine minimal competence in mechanics, even when they are sound, do not do more than touch the periphery of our courses. And we need to define as clearly as possible the objectives of our courses so that better testing programs can emerge. For reasons discussed in Section Five below, the nineteen institutions in our system have been able to come up with a far greater sense of agreement about objectives than has been possible nationwide.

3. STRENGTHS AND WEAKNESSES OF OBJECTIVE TESTING

The whole issue of objective testing is so complex, and so much research has been done on it, that to summarize the research risks superficialty and error. Most of the research on this question has been done by the Educational Testing Service, which has been giving English tests to large numbers of students since it was established in 1948. Several general conclusions emerge from the various reports produced by the highly capable scholars ETS has employed in this area:

1) Only those who know little about testing have unlimited faith in test scores; the specialists are well aware of the limitations and fallibility of any kind of test, and

2) the best test in composition will combine the most reliable elements of both objective and essay testing.

All evidence shows that both kinds of tests have important strengths and serious weaknesses; it is important to state here that there is no necessary conflict between essay and objective tests. We would, in fact, argue strongly against any equivalency testing in freshman English that did not include both.

Here are five conclusions that we support in the area of objective testing in freshman English:

A. Most of the objective tests available are poor, some scandalously so. We should not succumb to the feeling that people who make up tests must know what is going on in the field of English; many of them don't. Anyone with knowledge of modern linguistics or dialectology, for instance, would find some of the routine questions about "correctness" or the locating of supposed errors quite absurd. As one reads through test after test, he becomes convinced that the principal skill tested, repeatedly, is the ability to take tests, that is, the ability to discern the point of view of the test maker, and hence to guess shrewdly the "right" answer. No wonder the results on such tests correlate nicely with success in school, which is, after all, normally based on the same skill.

In short, the well-known deficiencies of multiple-choice testing still weaken most such tests. Here, for one example, is a question from one of the newest and most popular tests in English composition (slightly changed to avoid copyright difficulties); it illustrates the typical bad question still being written:

> English speaking musicians use professionally large numbers of words from which one of the following languages?
>
> a. German
> b. French
> c. Spanish
> d. Latin
> e. Italian

The test makers are obviously looking in this question for a scrap of information about the ways in which English uses foreign

words, in this case the Italian vocabulary for some aspects of musical notation. Some students may in fact pick up such information in a composition course, though it seems unlikely; but the student most able to fill in the proper square is likely to be the one whose parents wanted to and could afford to give him music lessons as a child. Those not so privileged (including, no doubt, some fine musicians) are not likely to know the answer, regardless of their writing ability. And someone who knew too much--say a specialist in medieval music--might even give the "wrong" answer, Latin.

At the same time, a few objective tests are noticeably better than the rest, and we ought to guard against uninformed judgements about all objective testing. Sometimes committees responsible for developing a test are wholly informed and up-to-date--sometimes, indeed, they are leaders in the field--and the test itself is sometimes reviewed with such elaborate care that the routine problems of objective testing are largely or wholly removed.

B. Writing ability is a highly complex combination of many skills; objective tests measure some skills analogous to and involved in writing, but can not measure all such skills and hence can never be wholly valid. We ought to distrust any objective test that claims to test writing ability in its entirety, and we should inquire suspiciously into the validity of such claims. On the other hand, there are skills which are closely connected with writing ability (for example, size and accuracy of vocabulary, or reading comprehension) which can be measured objectively with a high degree of reliability. We can and should demand that any objective test

we use examine particular skills with demonstrated validity, that it be free from the obvious flaws of such tests, such as social bias and ambiguity, and that it not advertise itself as testing more than it in fact does test.

C. Within some important limitations, objective testing can be a highly reliable and economical method of measuring achievement. Dr. Paul Diederich, Senior Research Associate at ETS, and one of the most experienced scholars in the country on English testing, writes that he usually expects, when measuring a single test against a reliable series of writing evaluations, "a correlation of about .65 with a good reading test, .55 with an objective test of writing skills, and .45 with grades on an essay given by trained readers under close supervision." These are discouraging figures: a correlation below .30 approaches irrelevance; professional designers of objective tests aim for .90 and are distinctly unhappy below .80. But we must recognize the fact, demonstrated repeatedly, that one good objective test will correlate more highly with a student's writing ability (using a series of writing samples as a base) than will one good essay test. This is a convincing argument that the equivalency test we approve should contain an objective section.

D. Since objective tests do not test writing ability directly, but only a few skills that are part of or associated with it, no objective test should be used alone as a measure of writing ability. It is essential that an essay be part of any writing equivalency test that seeks to measure college-level skill.

Everyone, even the most avid defender of objective testing, knows that some students can do well, or at least passably, on objective tests in writing and yet write abominably. In addition, whenever impersonal testing occurs, there are bound to be occasional instances of cheating, impersonation, and other outrages endemic to a test-oriented society. For these very practical reasons, essay tests are needed to increase the validity and security of the whole testing process.

E. Every English teacher's experience that writing ability is closely akin to reading ability is borne out by correlation studies. (Note that Paul Diederich, as cited above, expects a higher correlation with writing ability from an objective reading test than from an objective writing test. ETS reports tend to confirm his expectation.) This finding supports the common practice of spending much time in freshman English on careful analytic reading of all kinds of writing, including, but not restricted to, imaginative literature. Capable writers are almost always capable readers, and it is reasonable to expect that careful training in reading will help the development of writing ability. Since writing and reading are generally linked in the course work, and improved reading ability is a normal objective of freshman English, a test designed to give college credit in the course must include a substantial reading component. It appears possible to test reading ability with some accuracy by objective examination, and it appears possible to test general reading ability at least in part by the use of a valid and reliable general literature examination. But we must be careful that any reading

test we use is college-level and substantive. It is simpler to ask for the correct spelling of Shakespeare's name (though Shakespeare himself would not know) than to obtain and evaluate a response to King Lear's changing relationship to his daughter Cordelia.

4. STRENGTHS AND WEAKNESSES OF ESSAY TESTING

We ought to have no illusions about the reliability of essay testing. To be sure, it is the criterion of writing ability; it is the only way to see the real thing. Nonetheless, such tests have many important problems of which we need to be aware.

Perhaps the most significant problem for the reliability of essay tests is the large difference in quality of the essays written by a single student. Yesterday's paper is noticeably worse, or better, than today's and, of today's papers, the one on topic A is far superior to that on topic F. An essay does not measure writing ability as an abstract quality, but a student's ability to write on a certain topic on a certain day under test conditions. It is certainly conceivable that the student whose failing paper you may have read last night could have handed in his paper with a bored sigh of relief, and gone home to write his girlfriend in Cucamonga a witty, intelligent, mechanically accurate analysis of the test he had suffered through and of the agonies of the professor who would have to evaluate it.

The second most important problem is the difficulty in achieving reliable grading of essay tests. Even under the most carefully controlled and supervised reading conditions, it is hard to find readers who agree consistently about the quality of given essays. And the studies analyzing results under more usual circumstances, when students are writing on different topics, and when we know the identity of the writers, are really depressing.

But it is possible to establish testing and grading conditions to bring the reliability of essay testing to a useful point. It is clear that, as the ETS publication The Measurement of Writing Ability (1966) states, "The combination of objective items (which measure accurately some skills involved in writing) with an essay (which measures directly, if somewhat less accurately, the writing itself) proved to be more valid than either type of item alone."

Finally, it is educationally necessary to require a student to write during any test of writing. We need to validate objective testing by guarding against students who may have learned to perform well on tests, but who cannot write competently. Suppose we were to choose a simple, well-constructed spelling test as the equivalency examination (we won't, of course). The first time we used it, the results might well be acceptable; most (but certainly not all) good writers happen to be pretty good spellers. But the next time, those students who did not "waste" their school years writing, but instead studied spelling, would greatly improve their scores. In time, the exam might well stimulate mindless cramming and devalue the writing act itself. This would be the effect whatever combination of skills a non-essay test might examine; unless we include an essay test in our examination, we run the danger of defining writing as not-writing, and this would be a position without validity or integrity.

5. EQUIVALENCY TESTING IN FRESHMAN ENGLISH IN THE CALIFORNIA STATE UNIVERSITY AND COLLEGE SYSTEM

Conditions are favorable for the development of a responsible and accessible equivalency test in freshman English within the California State University and College system. Not only is the administration of the system on record as urging such testing in general, but the English Council itself has endorsed it in principle. In addition, some of the

problems we have listed in relation to freshman English courses and to testing in these courses are much more easily resolved within the system than on a nationwide basis.

For example, the contributions of the English Council to communication among the college departments has led to some general agreement about the objectives of freshman English in our institutions. Again, for various reasons, the English departments of the State University and Colleges have tended to devote a substantial portion of their time and some of their best energies to the development of freshman English. Hence, the nationwide neglect and fragmentation of such courses has not been a major matter here; indeed creative experimentation, innovation, and the development of new materials in such courses have marked our recent history.

The relative ease of communication among the 19 institutions, the general seriousness and spirit of innovation with which we approach the course, and the substantial size of our combined student bodies all argue for the possibility of a well-planned and appropriately financed examination that could have nationwide implications. Indeed, the importance of what we are here undertaking has not escaped the notice of ETS and CEEB; they have given strong assurances that the two organizations will bring their resources, experience, and knowledge to help us accomplish aims so consistent with their public position on credit-by-examination. The College-Level Examination Program has run into some important opposition from faculties, most particularly faculties in English and mathematics, numbers of whom have found the general examinations in these areas unacceptable. In response, ETS and CEEB have recommended various uses of subject examinations in these areas and are developing new examinations in both fields. Those responsibile for CLEP are determined to regain the confidence of these faculties.

We stand to benefit from a strong working relationship with ETS, which has done most of the valuable research in testing in our field, since this accumulated expertise (though by no means infallible) is an invaluable resource.

The testing program we recommend has four features to it, each of which is discussed below: A) A coherent statement of the aims and objectives of freshman English, B) A test, including both objective and essay parts, which is demonstrably responsive to these aims, calls for an appropriate college level of proficiency, and is valid and reliable, C) Administration of the test reliably and professionally, and D) Professional and sensitive use of test results. Such a program is not only academically sound, but financially and technically practical; we propose it go into operation for the fall of 1973, with initial testing to begin as early as spring 1973.

A. Aims and objectives of freshman English

Freshman English calls for development of reading and writing ability--including the effective uses of reference and resource materials--as well as the acquisition of knowledge about the English language. A student should demonstrate the college-level ability

1. to recognize and use appropriate language (rather than merely to classify "errors"),
2. to recognize and use the basic processes of clear thought and clear communication, and
3. to read expository and imaginative writing with understanding.

B. The Test: Objective and Essay

The test should contain both essay and objective parts. Six semester units of lower division credit, or its equivalent,

should be given for successful completion of an examination of 3 hours, consisting of 90 minutes of objective testing and 90 minutes of a carefully designed essay test.

C. Administration of the test

1. Proposed new CLEP Freshman English Test: Fall, 1974

We have great hopes that the proposed new CLEP Freshman English Subject Examination will be satisfactory for our purposes. We have confidence in the committee of examiners devising the test (Richard Braddock, University of Iowa; Greg Cowan, Forest Park Community College, Missouri; Marianne Davis, Benedict College, South Carolina; and Walker Gibson, University of Massachusetts) and respect the committee's statements about what it is seeking to accomplish. In addition, we have examined six 45-minute pretests containing approximately 450 questions written by college English teachers to the specifications of that committee. These pretests constitute an item pool from which about 200 questions will be drawn to yield two editions of 90-minute CLEP multiple choice subject examinations. On the following page are the test specifications developed by the committee of examiners. The questions on the pretests seem specifically designed to avoid the usual faults of short-answer testing, and seem generally to examine the kinds of skills we have agreed are among our most important objectives.

In addition the new CLEP freshman English test includes a 90-minute optional essay section which we can and should require. The committee preferred a required essay section

CONTENT / SKILLS	I. 40 M/C Q. Reading for Meaning (Exposition, Poetry, Prose)		II. 40 M/C Q. Manipulation of Language (Diction, Semantics, Syntax, etc.)		III. 40 M/C Q. Editing and Writing Ability Library and Research (Methods - Modes of Writing)		TOTAL
A. Knowledge	10	Kinds of organization and development - logical, etc. Facts about form.	10	Sentence sense and forms, functions of words in contexts.	5	Dictionary and Research skills. Reference sources. Language styles.	25
B. Comprehension	10	Basic content - Understanding ideas and implications - themes and facts.	10	Meaning of locutions and utterances, grasping analogies, comps. and contrasts, etc.	5	Organization, and Development of paragraphs, meaning and point of view.	25
C. Application			5	Reasoning from the specific to the general - conslusions and deductions.	15	Unity, Coherence and Emphasis. Use of editorial concepts applied to sample writing.	20
D. Analysis	10	Tone, figures of speech, identification of voice, purpose and audience.	5	Identification of relationships of words and phrases.	15	Tone-audience-suitability of language to purpose - grammar and usage.	30
E. Synthesis	10	Judgments about effectiveness, style and form.	10	Making inferences and judgments about form and meaning of words in contexts.			20
	6 passages, 2 each (Exp., Prose, Poetry) Contemporary, relevant materials 6 to 8 items per passage 100 - 300 words/passage		Majority discrete questions - Some sets of 2 or 3 on common stimulus - Model sentences, etc.		Simulated student's writing - 4 or 5 passages with 4-5 items each. Remainder - discrete items based on short stimulus.		120

as part of the test itself, but CLEP's policy is to let the decision on requiring the essay rest with the institutional score recipient. Everyone involved in creating the test agrees upon the value of the essay, however. Here is the policy of the CLEP program in relation to essay testing for the new CLEP Freshman English Test (exerpted, with permission, from an ETS memorandum dated July 14, 1972):

"The CLEP Program can offer a most positive alternative in the special case of this new freshman English test. This will permit and promote the careful, rational use of the optional essay section without penalizing those candidates whose essays would be misused or ignored:

"(1) If the committee makes its strongest possible recommendation urging recipient English departments to require the essay, the program will develop and distribute widely a special publication, aimed at college faculty members and departments, that will highlight the committee's recommendation. Colleges accross the country are in the process of developing policies of credit by examination through CLEP. A strong recommendation by the committee that this test is incomplete without a carefully prepared and graded essay should be, we think will be, welcomed by most recipients of scores. These schools can, should, and will in turn make it clear to individuals seeking credit that the essay <u>is required by the recipient institution</u>."

We expect to follow the development of this new test with keen interest, and are prepared to recommend its use if it fulfills its promise. We will seek to be included in the norming studies for objective portions of this test in the spring of 1973, and we will explore ways to conduct concomitant

norming of the optional essay section for students in our system. We have been assured by the Director of the CLEP program that the program will make tests available to us for these purposes at no charge, and will assist us in our validity studies. Unfortunately, while CLEP designs and provides for an optional essay, the receiving institutions must themselves provide for the grading of the essay question. Therefore, funding from the California State University and College budget will be needed in the 1972-73 fiscal year to establish an organization to read and evaluate essays for this test (or, indeed, for any other); this arrangement must be carefully and professionally set up, so as to assure the reliability and validity of the entire program. We expect to be able to draw upon California faculty experienced in AP and other organized essay grading efforts to assure the professional caliber of this essential operation; ETS specialists in this area stand ready to assist us.

However, because of the elaborate evaluation this new CLEP test will undergo, it will not be available for our use in September 1973. We thus need to choose an acceptable alternative for the year ahead, even as we watch the development of what may well be a CLEP test we can accept without qualms.

2. Analysis and Interpretation of Literature: Fall 1973

We recommend the following as a responsible short-term solution for the 1973-74 academic year only:

A three-hour examination consisting of the 90-minute objective CLEP Subject Examination, Analysis and Interpretation of Literature, and either its 90-minute essay section or one of our own devising.

The disadvantage of this short-term solution is that the test does not deal with composition aside from literature, and that no norms have been developed specifically for our student population.

The advantages of this proposal, however, are important:

a. The test exists, and has been well received throughout the country and within our system. It contains a highly reliable and valid objective test (according to the elaborate studies conducted by ETS), which will serve the necessary measurement function of the objective portion of our test.

b. The Literature test, while not ideally suited for all aspects of freshman English, is skewed in the direction of rigor rather than ease. It is a college level examination.

c. Reading skill correlates closely with writing skill, and this carefully constructed reading test, along with a 90-minute essay test, is more appropriate for our short-term use than any objective so-called composition test.

d. Two new, up-to-date, editions of this test will be available for our use in 1973. These new editions will improve an already impressive test.

e. CLEP has no objections to substituting an essay test of our own devising for the essays on literature now part of the examination. We can select appropriate essay questions for our purposes as the testing date

approaches, or we can accept those prepared by the CLEP committee (William Vesterman, Rutgers University; W. O. S. Sutherland, University of Texas; Mary Rion Hove, St. Olaf College) with the advice of the ETS test specialists.

3. Essay Grading: Supervision and Expenses

We resolve that the English Council will select a committee with continuing responsibility for supervision of the testing program. We need further reports on the development of the new CLEP Freshman English test, and since there is no national grading system for CLEP essays, we need to supervise the entire process of essay grading.

We propose that the English Council, funded through the Chancellor's Office, take responsibility for evaluating the student essays written for course equivalency in English. We can as a body ensure the integrity, consistency, and quality of essay grading far better than can any other office. Since essay grading is complex and expensive, it is bound to be vulnerable; under our direction it will be less assailable, less costly, and more reliable than any but a national system such as used by AP.

The cost of reliably grading large numbers of essays is not prohibitive (about $6.00 per exam, based on tentative estimates of costs shown in Appendix I), when measured against the potential savings for students and the system, and when placed against enhanced recruitment of able students, this expenditure in fact becomes a great bargain.

The cost of developing the examinations we recommend are being borne by CEEB; the cost of taking the examination is borne by the candidate seeking credit; the costs of scoring, reporting, and transcript service for the objective test are provided by ETS; the cost of scoring and using the essay section of the test needs to be provided by the State of California. During the 1972-73 fiscal year, this cost should, we suggest, be paid by the fund for innovative programs. But after the 1972-73 fiscal year, the faculty staffing formula should provide for the program, which obviously calls for continuing attention from the English Council and for maintaining a pool of trained readers. We hope that costs of grading can be reduced, as we gain experience; it may be that the scores on the objective test will be so valid for our purposes that papers of those on the upper and lower end of the scale will not need to be read.

See Appendix I for a tentative budget, tabulating the anticipated expense of grading 5,000 90-minute essay tests.

4. Passing Scores

We accept the recommendation of the independent Council on College-Level Examinations for the acceptable passing score on the objective part of the test. The Council recommends credit be granted for scores at or above the mean score for C students on the CLEP national norm. For the Analysis and Interpretation of Literature tests, that is a score of 49 or roughly the 50th percentile. (We may wish to use a California rather than a national mean score, when such local norming

takes place.) The essay test will need to be scaled by the chief reader and his assistants after the scoring has been done, and the two scores combined.

5. The Use of Test Scores

The use of test results requires careful attention and planning. Those who have passed the test and received credit for the college course work should be fully informed of the value of what they have achieved in academic and developmental terms--not merely mechanical or financial language; they should be urged to take more advanced work in English in order to develop their capacities further. Thus the placement value of this kind of testing should be exploited, even if course equivalency is the major purpose.

The individual colleges and universities should also retain flexibility in the use of test results, even if credit is granted systemwide. A student who does not succeed in passing the equivalency examination may wish to apply for a challenge examination at his own institution; he should have the opportunity to do so, if the institution wishes to continue offering such tests on a local basis.

The student should have the option of how he will use credit gained by examination. The experience of AP students is illustrative in this regard; these students, with their head start, take more college units than do students without AP credit. Certainly, careful and sensitive counseling, advisement, and

guidance are essential to this program, and not only for those likely to be successful in it. Those with little chance of success ought not to be encouraged to take tests covering college-level work they do not know; those succeeding at the tests should be encouraged and guided in their self-motivated quest for learning. Decisions, however, must always rest with the student, and each institution should seek to develop appropriate ways to help the student decide wisely.

6. The Colleges and the Schools

Since it is not to be expected that most, or even many, high school graduates will in fact have accomplished college-level work in English, without special training, no equivalency test program is complete without close liaison between the colleges granting credit and the schools. For a college-level equivalency program to succeed for more than a few individuals with unusual training or talents, the high schools will need help and support in providing formal college-level opportunities for all students who may profit from such opportunities. Such an innovative approach requires not only subject field communication between the colleges and the schools, but also a deliberate program of action on the part of the Chancellor's Office and the State Department of Education. We urge those agencies to initiate and foster a large-scale effort to assist the schools in establishing appropriate curricular offerings, so that the equivalency program we recommend can in fact be open to all potentially qualified students.

6. HOW EQUIVALENT IS EQUIVALENCY?

Even as we endorse equivalency examinations and proceed in all good faith to administer them, we need to reassert the value of our freshman English programs. After all, only a small percentage of our entering freshmen are likely to have learned college level skills in our field, and even some of those receiving credit may well seek to take freshman English in order to receive the less measurable benefits of the course.

Freshman English, as well as many other college courses, offers various experiences that have little to do with measurable skills, and yet that can be of great educational value to students. For example, as Thomas Wilcox puts it, "The English class may offer the freshman his only opportunity to participate in the free exchange of ideas and confer with a professional intellectual. This may be the best reason for limiting the size of freshman English classes and, indeed, the chief justification of freshman English itself." At a time when humanizing higher education has become much more than a slogan, we should not overlook the humanizing effect of a good freshman English course. "Students often testify, as they look back, that their freshman English course first brought their minds to life.... Because freshman English classes are still relatively small in most institutions, the instructor is often able to provide individual help for the student; he often becomes a counselor as well as a teacher, just because he is less remote than the lecturer in the large introductory courses." (Robert Gorrell, "Freshman Composition," The College Teaching of English, ed. John Gerber [New York,] 1965, p. 92)

If equivalency becomes one more mechanical device to turn education into processing, we will have done our students and our society a significant disservice, even if we have saved them some cash.

If equivalency becomes a simple matter of certifying minimal competency, without a concomitant push for achievement of individual excellence, we will have denied our mission.

We need to hold fast to our purpose as educators of individual students, even as we must get involved in the machinery of testing for units. The surest way for us to keep equivalency testing to its stated purpose of fostering and individualizing education in our field is for us to supervise directly a responsible professional program such as the one we here propose. Our aim, after all, is to help students educate themselves; we should expect that students will continue to come to us for the best we have to offer, and we can certify their achievements in various ways. Equivalency test scores may well be equivalent to our course grades, but the full and rich experience of language and literature, however measured, has no equivalency.

EXPENSE OF ESSAY READING

Following are estimates of the expense of reading with reasonable reliability 5,000 90-minute examinations, each composed of three separate essay questions. These estimates assume the following:

a. Three independent readings will be given each paper (one reading for each question).
b. Five minutes of reading time will be required to score each essay or a total of 15 minutes for each test.
c. Six tables of eight readers and one table leader each will be required for the reading; two tables for each question.
d. Each reader will receive an honorarium of $300 for 4½ days of work; each table leader will receive $350 for 5 days work.
e. An experienced chief reader will organize and direct the reading; $1,000 should cover his honorarium, travel, and expenses.

Honoraria (48 readers @ $300 and 6 table leaders @ $350)	$16,500
Per Diem (Housing and meals @ $22 per day)	5,412
Transportation (Average $50)	2,700
Chief Reader (Honorarium and expenses)	1,000
Clerical and Data Processing (Combining the 3 separate scores; combining the total essay and objective test scores; weighting scores appropriately, etc.)	5,000
TOTAL	$30,612

OBJECTIVE TESTS IN FRESHMAN ENGLISH

The following objective tests were made available by publishers for examination by the writers of this report. The College Proficiency Examination Program (CPEP) Examinations created and used by the University of the State of New York, were not made available; there are, no doubt, other tests in use, or *in potentia*, that we have not seen. We did, however, attempt to examine every widely available test designed for freshman English.

American Guidance Service, Circle Pines, Minnesota

- Essentials of English Test (forms A and B), by Dora V. Smith and Constance M. McCullough, rev. 1961 by Carolyn Greene

Bobbs-Merrill, New York

- Analytic Survey Test in English Fundamentals (form 4), by J. Helen Campbell and Walter Scribner Guiler

Bureau of Educational Measurements, Emporia, Kansas

- Barrett-Ryan English Test (forms I, II, III, VI, 1948, 1954)
- Barrett-Ryan-Schrammel English Test (forms EM, DM)
- Hoskins-Sanders Literature Test (forms A, B)
- Walton-Sanders English Test (Test I, form B; Test II, forms A, B)

Educational Testing Service, Princeton, New Jersey

- CLEP General Examination, English Composition
- CLEP Subject Examinations
  - English Composition
  - American Literature
  - English Literature
  - Analysis and Interpretation of Literature (six forms)
  - Freshman English (six pretests)
- Undergraduate Program (UP)
  - Literature Test
  - European and American Literature Test (modular complement to the Literature Test)
- Cooperative English Tests (forms 1A and PM)

Harcourt, Brace and World; New York

- Missouri College English Test, by Robert Callis and Willoughby Johnson (form B)

Houghton, Mifflin, Co.; Boston, Massachusetts

The New Purdue Placement Test in English (forms D and E), by G. S. Wykoff, J. H. McKee, and H. H. Remmers

McGraw-Hill, Monterey, California

Test of English Usage (form A), by Henry D. Rinsland, Raymond W. Pence, Betty Beck and Roland Beck

Educational Skills Tests, College Edition: English (form A)

Psychometric Affiliates, Brookport, Illinois

College English Test (forms A and B), by A. C. Jordan

LIST OF CORRESPONDENTS

The authors of this report are particularly grateful to the following specialists in the field of testing in English, for their detailed and valuable letters. If the report is valuable to the profession as a whole, it will be in large part due to the participation of the profession as a whole.

G. C. Allen, Deputy Director, School of Education, University of Sussex, Brighton, Sussex, England.

Richard Braddock, Coordinator, Rhetoric Program, University of Iowa and Editor of Research in the Teaching of English, Iowa City, Iowa.

J. N. Britton, Goldsmith Professor of Education, University of London, England.

Rex Brown, Assistant to the Director, Research and Analysis, National Assessment of Educational Progress, Denver, Colorado.

Jonathan Bryan, Associate Professor of English, Northern Virginia Community College, Annandale, Virginia.

Michael Cartwright, Coordinator of Freshman English, California State College, Bakersfield, and Director of the Association of Departments of English Survey of Freshman English Programs.

Paul B. Diederich, Senior Research Associate, Educational Testing Service, Princeton, New Jersey.

Gerhard Friedrich, State University Dean, Academic Program and Resource Planning, Office of the Chancellor, The California State University and Colleges.

Walker Gibson, Professor of English, University of Massachusetts, Amherst, Massachusetts.

William F. Irmscher, Director of Freshman English, University of Washington, Seattle, Washington.

Arnold Lazarus, Professor of English, Purdue University, Lafayette, Indiana.

William J. Leary, Professor of English, California State University, Los Angeles, California

Stoddard Malarkey, Director of Composition, University of Oregon, Eugene, Oregon.

F. T. Naylor, Director for Advanced Level Examinations, Schools Council, London, England.

Alan Purves, Professor of English, University of Illinois, Urbana-Champaign, Illinois.

J. Stephen Sherwin, Chairman, Freshman Program, SUNY College at Buffalo, New York.

Blanche J. Skrunick, Acting Director, Basic Writing Program, The City College of the City University of New York.

W. O. S. Sutherland, Professor of English, University of Texas, Austin, Texas.

TO: Dr. Robert Bess, Director
Academic Projects
Office of the Chancellor

DATE: December 4, 1973

FROM: R. W. Lid and Edward M. White
English Equivalency Testing Program

Attached is the full budget for the English Equivalency Testing Program, including funds already allocated and funds requested. Thus far we have been granted $15,741.00 start-up costs. We are requesting an additional $60,313.00 for a total of $76,054.00

For convenience the budget has been itemized under three separate parts:

(1) Start-up costs, allocated and requested
(2) Reading costs requested
(3) Follow-up costs requested.

The third category runs beyond the current fiscal year.

As to the request for additional funds under item (1) above, the original budget was for the Fall semester only. Within that budget we chose to economize by stretching faculty assigned time through a second quarter. Hence the requested additional reassigned time is for one quarter-course only, together with an adjustment figure for the semester/quarter differential. Beyond the above, there are some added clerical and operating costs which occur as we approach the reading period. Also, the State Printer's estimate for the printing of the brochure is running higher than anticipated, probably because of rising paper costs.

The Reading Conference costs are all based on 6,000 candidates taking the test. Should there be a substantially larger number, we would have to request additional funds--unless policy were set in advance that we close registration at some point. Conference costs are figured at a substantially lower figure than last year based on workload and production figures for Summer 1973. In effect, we will be reading half again as many test booklets (6,000 instead of 4,000) for only a marginal increase in costs (3-4 %). If we were to go over the estimated 6,000 candidates, the additional cost would be five dollars per test booklet to 7,500. It is of course assumed that if we come in under the 6,000, we will hire proportionally less readers. Finally, it should be pointed out that transportation and per diem have probably been figured high; we have been cautious because the gas shortage may mean that fewer of our readers will choose to commute this year.

As to Follow-up costs, they represent Dr. White's seeing the final processing and mailing of results to their conclusion. They also include the writing of a detailed final report and the making of a statistical study.

Two final points: (1) Last year Dr. Smart offered the Deans of Admission postage costs for mailing the brochures. Long Beach billed us $245.00. Had others done the same, we would have quickly gone over our budget. We would like a policy for this year that excludes such billing; if not, then we would like additional funds granted for each such charge made against us. (2) The only new item in this budget over last year's is computer costs. We have been unable to ascertain them precisely as of this date and the figure given is an estimate.

The budget has been itemized in some detail. Should you have any questions or need additional data, please feel free to ask Dr. Lid for such information.

RWL/mr

English Equivalency Testing Program

Budget Summary

Start-up costs.....	$21,515.00
Reading costs.....	49,348.00
Follow-up costs....	5,191.00
	$76,054.00

Start-up Costs

Original Grant - Personal Services

Faculty Man Years: .50	$ 6,252
Temporary Help	3,415
Staff Benefits (12%)	1,018
Salary Savings	(194)
Total Personal Services	$ 10,491

Operating Expenses

Brochures	$ 1,700
Mailing	1,500
Supplies	500
Facility Reservations	100
Word Processing	200
Travel	1,250
Total Operating Expenses	$ 5,250
[1]Total Budget	$ 15,741

Additional Costs - Personal Services

[2]Faculty Man Years	$ 2,432
Temporary Help	1,000
Staff Benefits	292
Total Personal Services	$ 3,724

Operating Expenses

Brochures	$ 350
Mailing	200
Dittoing	200
Xeroxing	300
Travel	1,000
Total Operating Expenses	$ 2,050
Total Additional Request	$ 5,774

[1]This is the original budget (Ref. No. A73-217) of $15,004 as adjusted to include 7.5 salary increases requested/received by Northridge.

[2]This sum is needed to cover disparity between quarter and semester accounting systems ($1,042) and one additional course off for Dr. White ($1,390). Dr. White is to have 4/9 ths reassigned time (4 courses); Dr. Lid, .25 (2 courses) for AY.

Reading Costs

Summary

Reading Costs

Readers and Professional Team	$ 27,300
Aides	2,301
Per diem/Readers and Professional Team	9,850
Transportation/Readers & Professional Team	5,850
Lunches/Aides	247
Registration Fee Kellogg-West	1,500
Computer costs	2,000
Xeroxing	250
Telephone	50
Total	$ 49,348

Costs Itemized by Category

Readers and Professional Team

60 readers @ $300 (4½ days)	$ 18,000
10 table leaders @ $400 (5 days)	4,000
2 question leaders @ $500 (5 days)	1,000
1 chief reader (12 days)	1,200
1 associate chief reader (17 days)	1,700
1 secretary (6 days)	300
*1 statistician (6 days)	600
2 chief aides @ $250 (5 days)	500
78	$ 27,300

Aides

10 @ $2.00 per hr., 9 hr. day, 5 days	$ 900
5 @ $2.50 per hr., 9 hr. day, 6 days	675
3 @ $3.00 per hr., 9 hr. day, 6 days	486
4 @ $2.00 per hr., 6 hr. day, 5 days	240
22	$ 2,301

*For 1974 reading ETS will likely provide statistician and travel costs; we will provide per diem. Hence this particular sum will most likely not be expended.

Per diem/Readers and Professional Team

74 @ $25 per day for 5 days	$ 9,250
4 @ $25 per day for 6 days	600
78	$ 9,850

Transportation/Readers and Professional Team

78 @ $75 average	$ 5,850

Lunches/Aides

14 for 5 days @ $2.50 per day	$ 175
8 for 6 days @ $2.50 per day	72
22	$ 247

Registration Fee - Kellogg-West

78 Readers and Professional Team & 22 Aides (100 x $15.00)	$ 1,500
Computer	$ 2,000
Xeroxing	$ 250
Telephone	$ 50

Follow-up Costs - Summer 1974

Director's Salary (2 months)	$ 3,742
Temporary Help	1,000
Benefits	449
	$ 5,191

The California State University
English Equivalency Test

STATISTICAL PROCEDURES

May, 1973

William Cowell

I. The Test and Scores:

The California State University English Equivalency Test consists of two ninety minute parts; an objective part and an essay part. The objective part is the Subject Examination "Analysis and Interpretation of Literature" of the College-Level Examination Program (CLEP). The essay part has been prepared by the English professors of the California State Universities and Colleges.

Educational Testing Service will score the objective part and report scores on a scale of 20 to 80. The essay part consists of two questions. Each of the two essays will be read by two readers. Readers will assign grads on a scale of 1 to 6 (zero will denote that the student did not attempt an answer). The four essay grades will be added to obtain the essay total score. Essay totals can range from 0 to 24. Essay totals of 0-3 are possible only if the student does not attempt one or both of the questions.

A Composite score for the whole examination (objective and essay) will be determined by transforming the essay total score to the same scale as the objective scores and adding the objective and the transformed essay scores. The technique for transforming the essay scores is described below (Section III).

II. Candidate Groups and Norms Samples:

1) Total Group:

The candidates (potential California State University and College freshman) who register for and take the English Equivalency Test.

2) Equating Sample:

A sample of the Total Group. Since it is desirable to begin the work of combining scores before the essay reading is finished, the transformation of the essay scores to the CLEP will be based on a sample of the total group. This sample

will consist of the first 500 papers from the total group to have the reading completed. Since the papers of the total group will be scrambled before the reading begins, the sample should be representive of the total group. The representativeness can be checked by comparing the objective scores of the sample and total group. If the mean or standard deviation of the scores for the sample differ significantly from those of the total group, the sample should be increased until it is representative. This sample will be referred to as "The Equating Sample".

3) CSUC Norms Sample:

The sample of CSUC students who take the essay part of the examination as part of their English course. This group will be used in determining minimum passing scores.

4) CLEP Norms Sample:

The nation-wide sample of college students who participated in the 1964 norming administration of the CLEP Analysis and Interpretation of Literature examination.

5) Texas Norms Sample:

The sample of students who participated in the 1970 validity study of the CLEP examination at the University of Texas.

III. Equating of Objective and Essay Scores:

Equating is a method of transforming the essay scores to the CLEP scale to permit addition of objective and essay scores in a meaningful way. Equating may also be of value in determining minimum passing points.

Since both objective and essay scores will be available for the Equating Sample (see II above), it is possible to establish a table showing the CLEP score (on a scale of 20 to 80) equivalent to each essay total score. The process of obtaining this table is called "score equating". Several acceptable methods are available[1]. The equipercentile (curvilinear) method seems to be most appropriate for establishing comparable objective and

---

[1] Angoff, W. H.; Scales, Norms, and Equivalent Scores in R. L. Thorndike (Ed.) "Educational Measurement". (2nd ed.) Washington, D. C.: American Council on Education, 1971, 563-590.

essay scores. The scaled-score corresponding to each essay total score will be that score in the distribution of objective (CLEP) scores such that the percentage of the group scoring below that CLEP score is equal to the percentage of the (same) group scoring below the given essay score. Step-by-step procedures are given in Appendix A.

IV. Minimum Passing Scores:

The CSUC Norms Sample (see II above) will provide information to determine the score on the essay part of the examination that corresponds to the passing point as defined in the announcement of the program "A Head-Start at College" - i.e., performance at the level of the average "C" student finishing the college English course.

The minimum passing score corresponding to the average "C" student's performance is to be determined in terms of the combined objective and essay scores. Setting a single cut-off point on the Composite score allows for freely compensating part-scores; i.e., a very poor performance on the essay part can be compensated for by a very good performance on the objective part, and vice versa. It has been decided that limits should be placed on this compensation. To pass the test a candidate must earn a Composite score equal to or greater than the minimum passing score (middle "C") and he must earn part scores (objective and essay) equal to or greater than certain minimum scores (to be determined).

As was pointed out at the April 18th meeting, it would be very desirable from a psychometric point-of-view to have both objective and essay scores available for the CSUC Norms Sample. In that case minimum passing scores could be derived directly for both parts and the total. Because of the high costs in testing time and in money, only the essay part will be administered to the CSUC Norms Sample. Minimum passing scores for the objective part and the total test can be derived from other data which will be available. Two alternatives exist for estimating the level of performance of an average "C" student on the objective part:

1) The equating method described in Section III establishes a relationship between objective and essay scores in a way that allows direct comparison. For example, an objective score of 49 represents the same level of performance as an essay score of 49. Therefore, the CSUC Norms Sample essay score distribution could be used to establish two "cutting-scores"; a) the score corresponding to the level of performance of an average "C" student which eventually would determine the minimum passing score on the total test, and b) the minimum acceptable score on the essay part. Since the objective scores represent the same level of performance as the corresponding essay scores, it seems reasonable to use the same cutting scores for the objective part as is determined for the essay part. Step-by-step procedures are given in Appendix B.

2) An alternative for estimating minimum passing points for the objective part involves the use of the CLEP Norms Sample and/or the Texas Norms Sample. Each of these groups consists of college students enrolled in "appropriate" English courses and each group took the CLEP objective test. Data are available to indicate the level of performance of an average "C" student. However, average "C" is defined in terms of grading standards at non-CSUC institutions and "appropriate" courses were selected by non-CSUC staff to match the description of the content of the CLEP examination. The CLEP Norms Sample consists of students from 17 colleges representing a wide range of types of institution and geographical locations. The table below shows the mean CLEP score for the various course grade categories for each norms group.

	NORMS SAMPLE	
	CLEP	TEXAS
Mean Score for:		
"A" students	61	61
"B" students	55	55
"C" students	49	49
"D" students	41	41
"F" students	40	50
Total Groups	50	52

The remarkable similarity of the Texas and National Data is supportive of the use of national norms data to establish minimum passing scores. However, in using national norms data to establish minimum passing scores for the objective part, it is assumed that collective standards for course grades among the 17 colleges are similar to standards for the CSUC and that the relationship of CLEP scores to course grades is similar in the two situations.

Having two methods available for estimating objective cut-off scores, one method can be used to check on the other, or modify the other if substantial differences occur. I recommend that the equating method be used to determine the minimum passing scores and the CLEP Norms data be used to check, or modify, the results. I think it is desirable to

derive the minimum passing scores directly from California data. Appendix B describes the procedure in detail.

V. Checks on Reading Reliability:

Three types of checking and control will be used with respect to reading reliability.

1) Table Leaders will continuously spot check the grading of readers at their table and Question Leaders will continuously monitor the grading of all the tables reading their question. These checks are informal in that no record will be kept but differences between the judgments of reader and table leader or question leader will be resolved on-the-spot. This procedure is intended to maintain grading standards over the duration of the reading.

2) When the grading is completed for a paper, the cover (carbon) strip is removed and the whole set of four grades is visible. The aide removing the cover will inspect the grades for consistency. The grades for first and second readings of the same essay ought to agree, within reasonable limits. There is no reason to expect grades for different essays to be the same, so no checking is needed between the grades for the first essay and grades for the second essay. For convenience in referring to each of the four grades, let:

Q11 denote the grade for Question 1, first reading

Q12 denote the grade for Question 1, second reading

Q21 denote the grade for Question 2, first reading

Q22 denote the grade for Question 2, second reading

Then Q11 and Q12 should be compared and Q21 and Q22 should be compared, but no checking is needed for other possible pairs.

The difference between grades for two readings that is to be accepted is fairly abitrary. It is probably not feasible to try to resolve differences of 1 or 2 points, but differences of 3 or more points probably ought to be resolved.

The procedure for resolving differences of 3 or more points should be established prior to the reading. It is suggested that the aide return the paper with discrepant grades to the Question Leader. The Question Leader will read the essay himself or assign it to a third reader to try to determine which grade is in inaccord with the standards. The Question Leader will then discuss the discrepancy with the reader and/or Table Leader to maintain consistency of grading standards

as well as to try to resolve the difference in grades. It seems inappropriate for anyone to change a grade without the consent of the reader who assigned it.

An alternate procedure would be to return the paper to the Question Leader, obtain a third reading, and use all three grades. This procedure would be faster but does not have the beneficial effects on maintaining grading standards and allows "incorrect" grades to enter into a student's total score. The essay total should be adjusted to include only 2/3 of the sum of the three grades, rounded to the nearest whole number, plus the grades for the other question to keep the essay total on the same scale as regular papers.

3) The third type of check on reading reliability involves obtaining correlations between grades for first and second readings. This work can be done by the statistical aide following the recording of each day's production.

The reading reliability for question 1 is the correlation between grades Q11 and Q12. The reading reliability for question 2 is the correlation between grades Q21 and Q22.

VI. Validity:

The college professors of the students in the CSUC Norms Sample have been asked to supply two measures of performance in the college English course for each student:

1) the letter-grade in the course

2) a categorical ranking in a forced distribution

Corelations between essay total scores and each of the two performance measures can be obtained as the data become available to obtain validity coeficients for the essay scores.

CSUC ENGLISH EQUIVALENCY TEST
STATISTICAL PROCEDURES

APPENDIX A

PROCEDURES FOR EQUIPERCENTILE EQUATING

May, 1973

William Cowell

1. Before the reading begins the essay booklets from the regular candidate group should be put in a random order. It is particularly important that the Equating Sample be a random sample of the total group. If a candidate used more than one booklet, the booklets should be securely fastened together to withstand many handlings during the reading.

2. About 500 essay booklets selected at random from the regular candidate group should be mixed with the booklets from the CSUC Norms Sample. These thousand (approx.) booklets should be read first to enable the statistical work of equating and norming to begin as early as possible.

3. As the reading is completed on each batch of booklets, the grade cover (carbon) strips are removed and the aide should check for consistency of grades between first and second readings. Grade discrepancies should be resolved as quickly as possible and the paper returned to the aide. Booklets for the CSUC Norms Sample should be separated from those for the Equating Sample. A special code for the Norms Sample will identify those booklets. (See Appendix B)

4. As the completed booklets are sorted, the information from the Equating Sample booklets should be recorded on the form:

ROSTER OF SCORE INFORMATION
CSUC ENGLISH EQUIVALENCY TEST
EQUATING SAMPLE

The columns headed "Converted" and "Composite" will be filled in at a subsequent step. The Objective Score will be copied onto this form from the alphabetical roster of objective score information supplied by ETS. The information can be recorded in the order that the booklets are received; it is not necessary to alphabetize this roster. If a candidate has an essay booklet but is not listed on the objective score roster, remove the booklet from the Equating Sample.

5. After the information has been recorded on the roster, the information should be checked by another person and red dots placed beside the Objective Score and Essay Total Score to indicate that these data have been checked and found to be correct. If a discreponcy is found, cross-out the incorrect figure and write the correct figure in red, then have it re-checked.

6. Sort the booklets into 25 piles depending on the Essay Total Score (0-24). The number of booklets in each pile will give the frequencies for the "Distribution of Essay Grades".

7. After all the booklets for the Equating Sample have been read and the information recorded on the roster, score distributions for the Objective Scores and Essay Total Score will be obtained. Use the forms "DISTRIBUTION OF OBJECTIVE SCORES" and "DISTRIBUTION OF ESSAY TOTAL SCORES" to make a tally of the scores. One person can read scores from the roster while another tallies the scores on the distribution. Then, for the objective scores, reverse the jobs and make another tally to check the first. For the essay scores, a count of the booklets in each pile can be used to check the tallies. (See step 6 above) It is very important that these distribtuions for the Equating Sample be accurate since they will determine grades for the total group.

8. When an accurate tally has been obtained, write the frequencies (counts) on another copy of the form. Then fill in the "Cumulative Frequency" column, accumulating frequencies from the bottom to the top. The figures in the "Frequency" and "Cumulative Frequency" columns should be the same for the lowest score, and the cumulative frequency for the highest score should be the total number of candidates in the sample.

9. Have the "Frequency" and "Cumulative Frequency" data checked.

10. Fill in the "Percent Below" column. The percent below for any given score is the cumulative frequency of the next lower score divided by the total number of candidates (rounded to the nearest tenth percent). For example, to find the percent below for a score of 20, find the cumulative frequency for score-point 19 and divide that number by the total number of candidates in the sample. Then more the decimal point two places to the right (equivalent to multiplying by 100) and round to the nearest tenth (e.g., 48.7).

11. Have the percent below calculations checked. Place a red dot beside the percents as they are checked to indicate they have been checked and found correct.

12. Plot the CLEP objective scores against their percent below values on arithmetic probability paper. Check the plotting, then ink in the points. Draw a smooth curve through the points using a French-curve. (If a computer program for curve-fitting can be accessed via the ETS remote terminal, the computer program may be used to check on the graphical procedure.)

13. For each Essay Total Score, the corresponding CLEP Scaled Score will be found by using the graph obtained in step 12. For each Essay Total Score,

    a) find the percent below corresponding to that score on the "Distribution of Essay Total Scores"

    b) locate the corresponding percent below along the horizontal axis of the graph

    c) read the corresponding CLEP score from the vertical axis (to the nearest integer value)

    d) record equivalent Essay Total and CLEP Scaled Scores on the "SCORE CONVERSION TABLE".

14. Have step 13 checked.

15. Record the converted scores on the "Roster of Score Information".

16. Add the objective and the converted essay score to obtain the Composite Scores.

17. Have steps 15 and 16 checked.

18. Tally the composite scores and obtain the "DISTRIBUTION OF COMPOSITE SCORES" as in steps 7-11.

19. Obtain the correlation coefficient between objective and converted essay scores (including means and standard deviations of each score distribution).

CSUC ENGLISH EQUIVALENCY TEST
STATISTICAL PROCEDURES

APPENDIX B

PROCEDURES FOR ESTABLISHING MINIMUM PASSING SCORES

May, 1973 William Cowell

1. The essay booklets from the CSUC Norms Sample should be mixed with about 500 booklets selected at random from the regular candidate group. These thousand (approx.) booklets should be read first.

2. As the reading is completed on each batch of booklets and the consistency of grading is checked, booklets for the CSUC Norms Sample should be separated from those for the Equating Sample. (See Appendix A, steps 1-3)

3. As the booklets are sorted, the information from the CSUC Norms Sample booklets should be recorded on the form:

ROSTER OF SCORE INFORMATION

CSUC NORMS SAMPLE

Course grades and rank index can be obtained from the rosters supplied by the college professors. Essay grades and Essay Total should be recorded even if course grade and/or rank index are not available. The column headed "Converted" will be filled in later. The information can be recorded on this roster in the order that the booklets are received; it is not necessary to alphabetize.

4. After the information has been recorded on the roster, the information should be checked by another person. Red dots should be placed beside the essay total score and the course grade to indicate that these data have been checked and found to be correct. If a discrepancy is found, cross out the incorrect figure in red, then have it re-checked.

5. Sort the booklets for the CSUC Norms Sample into six piles depending on the code for Course Grade (A, B, C, D, F, and No Grade Reported).

6. After all the booklets for the Norms Sample have been read and the information recorded on the roster, six distributions of essay total scores should be constructed; one distribution for each of the grade categories A, B, C, D, F, and No Grade Reported. Steps 7-11 of Appendix A describe the procedures.

7. After these six distributions have been obtained and checked, a distribution of essay total scores for the entire CSUC Norms group can be obtained by adding the frequencies across the six distributions (for each essay score).

8. Find the mean and standard deviation of each of the seven distributions.

9. The mean of the distribution of essay total scores for the grade category "C" is the basis of the minimum passing score. Use the graph constructed in step 12 of Appendix A to find the CLEP Scaled Score (Converted Score) corresponding to the mean "C" grade. Multiply this score by 2 and round to the nearest integer value. This is the minimum passing score, using Method 1 as described in "Statistical Procedures".

10. To compare the results of the two methods suggested in "Statistical Procedures" for establishing the minimum passing score, add 48.54 to the CLEP Scaled Score corresponding to the mean "C" grade. (The value 48.54 is the mean CLEP score for the CLEP National Norms Sample.)

11. The minimum passing scores derived by each of these methods (steps 9 and 10) should be reported to the Director of the Reading for final determination of the minimum passing composite score.

12. The minimum passing part-scores must also be determined. This score should be set to minimize the number of classification errors (A, B, and C students who fail and number of D and F students who pass). To determine this point for the essay scores, construct two additional frequency distributions; one for the combined A, B, and C course-grade groups and one for the combined D and F groups. (Only the frequency columns need be filled in.) Inspection of these frequency columns will suggest one or more possible cutting points. The score which minimizes the number of classification errors should be selected. The Director of the Reading will make the final decision on this minimum passing part-scores.

13. After the "Score Conversion Table" has been completed (see Appendix A), the Converted Essay Scores can be added to the "Roster of Score Information".

ROSTER OF SCORE INFORMATION
CSUC ENGLISH EQUIVALENCY TEST
EQUATING SAMPLE
May 1973

Name	Soc. Sec. No.	Obj. Score	Essay Grades						Composite
			Q11	Q12	Q21	Q22	Total	Converted	

ROSTER OF SCORE INFORMATION
CSUC NORMS SAMPLE
May, 1973

144.

Name	Soc. Sec. No.	Course Grade	Rank Index	Essay Grades					
				Q11	Q12	Q21	Q22	Total	Converted

DISTRIBUTION OF OBJECTIVE SCORES
EQUATING SAMPLE
May, 1973

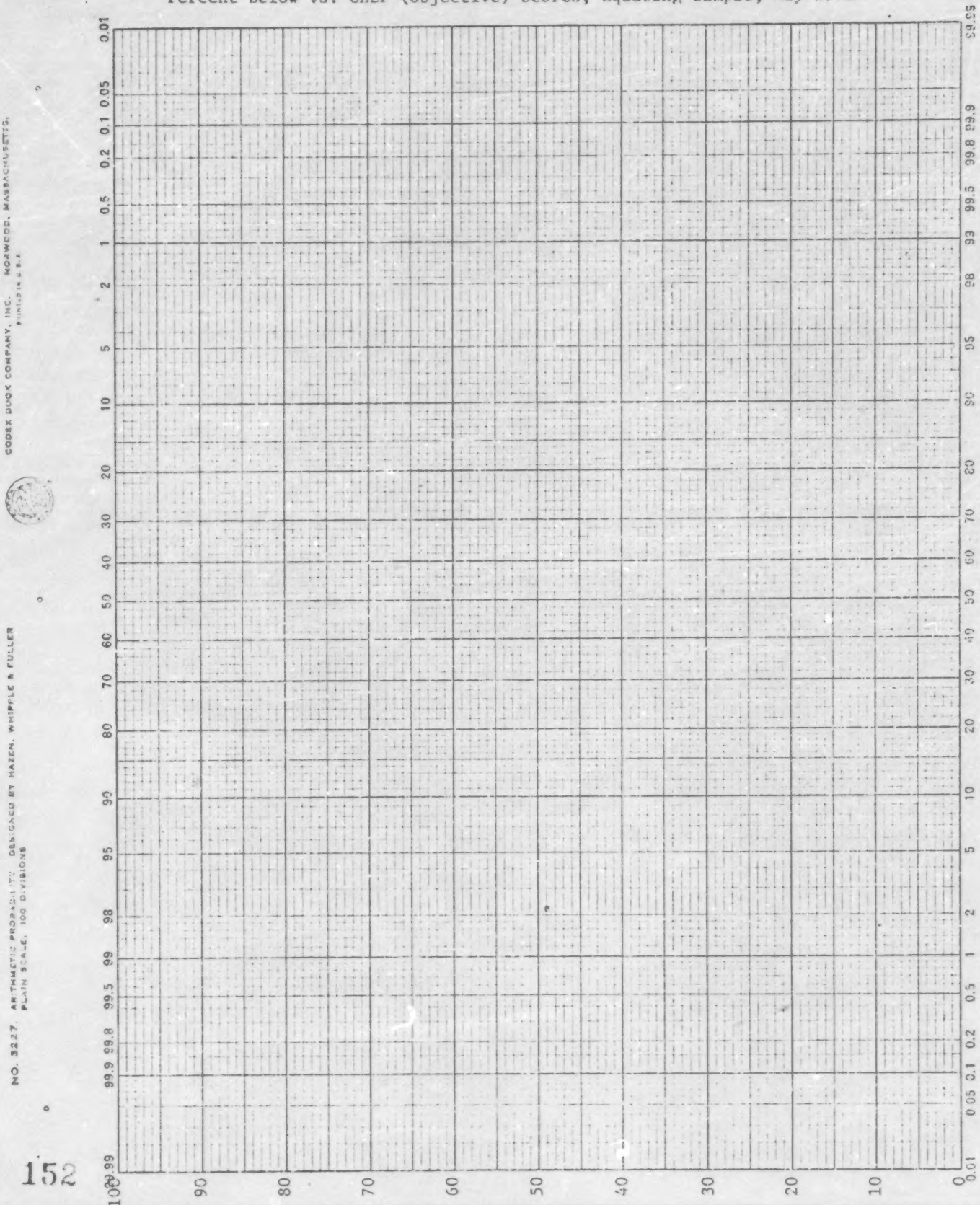
Objective Score	Frequency	Cumulative Frequency	Percent Below	Objective Score	Frequency	Cumulative Frequency	Percent Below
80							
79				49			
78				48			
77				47			
76				46			
75				45			
74				44			
73				43			
72				42			
71				41			
70				40			
69				39			
68				38			
67				37			
66				36			
65				35			
64				34			
63				33			
62				32			
61				31			
60				30			
59				29			
58				28			
57				27			
56				26			
55				25			
54				24			
53				23			
52				22			
51				21			
50				20			

Entered by: _____

Checked by: _____

California State University English Equivalency Test

Percent Below vs. CLEP (Objective) Scores, Equating Sample, May 1973



CODEX BOOK COMPANY, INC. NORWOOD, MASSACHUSETTS
PRINTED IN U.S.A.
NO. 3227. ARITHMETIC PROBABILITY, DESIGNED BY HAZEN, WHIPPLE & FULLER
PLAIN SCALE, 100 DIVISIONS

DISTRIBUTION OF ESSAY TOTAL SCORES
May, 1973

Essay Total Score	Frequency	Cumulative Frequency	Percent Below
24			
23			
22			
21			
20			
19			
18			
17			
16			
15			
14			
13			
12			
11			
10			
9			
8			
7			
6			
5			
4			
3			
2			
1			
0			

Entered by: _____

Checked by: _____

DISTRIBUTION OF COMPOSITE SCORES
CSUC ENGLISH EQUIVALENCY TEST
May, 1973

Composite Scores	Frequency	Cumulative Frequency	Percent Below
160			
155 - 159			
150 - 154			
145 - 149			
140 - 144			
135 - 139			
130 - 134			
125 - 129			
120 - 124			
115 - 119			
110 - 114			
105 - 109			
100 - 104			
95 - 99			
90 - 94			
85 - 89			
80 - 84			
75 - 79			
70 - 74			
65 - 69			
60 - 64			
55 - 59			
50 - 54			
45 - 49			
40 - 44			
35 - 39			
30 - 34			
25 - 29			
20 - 24			
15 - 19			
10 - 14			
5 - 9			
0 - 4			

Entered by: _____

Checked by: _____

SCORE CONVERSION TABLE
CSUC ENGLISH EQUIVALENCY TEST
May, 1973

ESSAY TOTAL SCORE	CONVERTED SCORE
24	
23	
22	
21	
20	
19	
18	
17	
16	
15	
14	
13	
12	
11	
10	
9	
8	
7	
6	
5	
4	
3	
2	
1	
0	

Entered by: _____

Checked by: _____

THE CALIFORNNIA STATE UNIVERSITY

English Equivalency Test

STATISTICAL REPORT

July, 1973

William Cowell

I. The Test.

The California State University English Equivalency Test was administered to 4,071 students on May 12, 1973. The test consists of two ninety-minute parts; an objective (multiple-choice) part and an essay part. The objective part is the Subject Examination "Analysis and Interpretation of Literature" of the College-Level Examination Program (CLEP). The essay part consists of two questions prepared and graded by the English professors of the California State University and Colleges (CSUC).

II. Scores.

(1) Objective Score. The Educational Testing Service scored the objective part and reported scores to the students and the CSUC English Test Center on the CLEP score-scale of 20 to 80.

(2) Essay Question Scores. Each of the two essays was graded by two readers who assigned scores on a scale of 1 to 6. Zero denotes that the student did not attempt an answer to the question.

(3) Raw Essay Total Score. The four essay question scores were added to obtain the raw essay total score. The essay total scores range from 0 to 24. Scores of 0 to 3 are possible only if the student did not answer one or both questions.

(4) Converted Essay Score. The raw essay total scores were transformed to the same scale as the objective scores to permit addition of objective and essay scores in a meaningful way. The technique for converting the scores is described below (Section IV.)

(5) Composite Scores. Composite scores for the whole examination were obtained by adding the Converted Essay Scores and the Objective Scores.

III. Candidate Groups and Norms Samples.

(1) Total Group. The total group consists of the 4,071 students who took the English Equivalency Test in May 1973.

(2) Equating Sample. The equating sample consists of the first 536 papers completed at the reading session. Essay books were scrambled before the reading to essentially randomize the order. The sample is not significantly different from the total group with respect to the mean and standard deviation of the objective scores.

	Objective Scores	
	Mean	S.D.
Total Group	47.45	8.85
Equating Sample	47.63	8.75

(3) CLEP Norms Sample. The CLEP Norms Sample consists of the 541 college students who participated in the 1964 norming administration of the CLEP Analysis and Interpretation of Literature Examination.

(4) Texas Norms Sample. The Texas Norms Sample consists of the 188 students who participated in the 1970 validity study of the CLEP examination at the University of Texas.

(5) CSUC Norms Sample. The essay part of the English Equivalency Test was administered to a sample of CSUC students completing a year of college English in an attempt to link performance on the test to performance in their course (as measured by their grade in the course). It had been planned to use the scores of the CSUC Norms Sample to establish the minimum passing score for the essay part of the examination. Since fewer than half the expected number of papers were received in time for the reading and because the scores were unreasonably low, it was decided to base the minimum passing scores on other criteria (see section VI, below). A summary of the score information for the Norms Sample is shown in Table 2. Although the data may not be very reliable, they do provide some indication of the relationship between college grades and grades on the essay part of the exam.

Table 2

Summary Data for CSUC Norms Sample
Raw Essay Total Scores

Course Grade	N	Mean	Standard Deviation
A	26	13.5	4.0
B	75	12.9	3.0
C	98	10.6	3.5
D	33	10.1	3.2
F	10	7.8	2.4
Total	242	11.5	3.7

IV. Equating the Essay and Objective Scores.

"Equating" scores means making them equivalent with respect to some characteristic. Here, the essay scores were transformed to the same scale (20 to 80) as the objective scores so that the transformed essay scores could be added to the objective scores to obtain a composite score in which the two parts would be equally weighted. Had

the raw essay scores been added to the objective scores, the essay part would have had a weight of only 1/4 of the total. Also, the transformation was done in such a way that the converted essay scores can be directly compared with objective scores. That is, a converted essay score of 48, for example, represents the same level of performance as an objective score of 48.

Of several acceptable methods of equating which are available[1], the equipercentile (curvilinear) method was selected as the most appropriate. The converted score corresponding to each raw essay score is that score in the distribution of objective (CLEP) scores such that the percent of the group scoring below that objective score is equal to the percent of the group scoring below the given essay score. For example, to find the converted score corresponding to a raw essay score of 15, we observe from the distribution of essay scores that about 68 percent of the group scored below 15 on the essay part. From the distribution of objective scores it is determined that a score of 53 is that score below which 68 percent of the group scored. So a raw score of 15 is equivalent to a scaled score of 53. Similarly, a scaled score is found for each raw essay score. Table 1 shows the converted score corresponding to each raw essay score.

V. Technical Notes about the Equating.

(1) The equating was done using data from the Equating Sample described in III(2) above. The sample was used instead of the total group so that the statistical work could proceed along with the essay reading. Score conversions and distributions could be prepared for the sample data in time for a meeting with Question Leaders at the reading. The time and cost of equating was also reduced by using a sample rather than the total group. Scrambling the essay booklets before the reading provided a sample which was representative of the total group.

(2) A method of graphical smoothing, described in "Educational Measurement"[2], was used to interpolate between percentile ranks in the distribution of objective scores.

(3) Although the equipercentile method was selected because it allows for a non-linear relationship between objective and essay scores, the correlation between raw essay and converted essay scores is .999. This extremely high coefficient of linear relationship indicates that the simpler linear methods[3] of equating could have been used with essentially the same results as the equipercentile method.

---

1. Angoff, W. H.; "Scales, Norms, and Equivalent Scores" in R. L. Thorndike (Ed.) Educational Measurement, 2nd Ed, Washington, D. C.: American Council on Education, 1971, 563-590.

2. Ibid., page 572.

3. Ibid., page 513.

TABLE I

SCORE CONVERSION TABLE
CSUC ENGLISH EQUIVALENCY TEST
May, 1973

Essay Total Score - Converted Score

24	76
23	73
22	71
21	69
20	67
19	64
18	62
17	59
16	56
15	53
14	51
13	48
12	46
11	43
10	41
9	38
8	35
7	33
6	31
5	28
4	26
3	24
2	22
1	20
0	18

VI. Minimum Passing Scores.

The minimum passing score for the total test was determined by setting a passing point for each part and adding the scores. For both the CLEP Norms Sample and the Texas Norms Sample, described in paragraphs III (3) and III (4) above, the average CLEP score for "C" students was 49. It was decided that 49 should be the minimum passing level for the objective part. The minimum passing score for the essay part was set at 14, which converts to a scaled score of 51. A composite score of 100 (49 objective + 51 essay) was thus determined to be the minimum passing score on the examination.

Setting a single cut-off point on the Composite Score allows for freely compensating part-scores; that is, a very poor performance

on the essay part can be compensated for by a very good performance on the objective part, and vice versa. It was decided that limits should be placed on this compensation. In addition to earning a composite score of 100 or more, a student had to achieve a minimum score on each part. These minima were set at 45 for the objective part and 48 for the essay part. (A converted score of 48 corresponds to a raw essay score of 13.) A student who scored 12 or lower on the essay part could not pass the exam no matter how high the objective score. A raw essay score of 12 represents the score level of uniform agreement among four readers that the essay performance is unsatisfactory. An objective score of 45 was set as the cut-off point because in both the CLEP Norms Sample and the Texas Norms Sample a score of 45 minimized the number of A, B, and C students who would fail and the number of D and F students who would pass.

VII. Number and Percent Passing.

Of the 4,071 students who took the test, about 33 percent (i.e. 1,343 students) passed and are eligible to receive six semester hours of college credit. From the distribution of scores on the objective part, it can be determined that about 43 percent (i.e. 1.733 students) "passed" the objective half of the test - that is, about 43 percent of the total group earned scores of 49 or higher on the objective part. Therefore, had only the objective part of the test been given and the same minimum score been used, an additional 10 percent would have passed. These students failed because their essay scores were unsatisfactory.

Corresponding data based on the distribution of essay scores are not yet available for the total group. Estimates based on data from the Equating Sample will be discussed in the following section.

VIII. Estimates Based on Data from the Equating Sample.

Estimates of the characteristics of the total group can be made from data available for the Equating Sample. It has been pointed out in paragraph III (2) that the mean and standard deviation of objective scores for the sample are very nearly equal to those of the total group. It is not clear, however, that the essay scores for the sample are representative of the total group. It is not uncommon, for example, to find that grading tends to become more severe near the end of an essay reading session. Until final essay results are tabulated, a direct check on essay data is not possible. As an indirect check on the similarity of the sample and total group, the percents passing the entire exam (objective plus essay) were compared. Whereas only 33.0 percent of the total group passed the entire exam, 37.5 percent of the Equating Sample passed. This rather large difference suggests that the following data be used with caution, allowing for the possibility of substantial changes when the final distributions are available. When actual data become available, these estimates should be disregarded.

(1) For the Equating Sample, 43.1 percent "passed" the essay part of the exam - that is, 43.1 percent of the sample earned scores of 51 or higher on the essay part. The difference between 43.1 percent "passing" the essay part and 37.5 percent passing the entire exam (i.e. 5.6 percent) is the percent of the sample who failed because of an unsatisfactory performance on the objective part. Had only the essay part of the exam been given and the same minimum passing score (i.e. 51) been used, an additional 5.6 percent of the sample would have passed. This estimate is probably a fairly reliable estimate for the total group because the sample and total group are very nearly equal on objective-part performance. Projected to the total group of 4,071 students, an additional 225 to 265 students would have passed if only the essay part been given.

(2) Of the 536 students in the Equating Sample, 218 had Composite Scores of 100 or higher, i.e. they had "passing" Composite Scores. Of these 218 students, 201 passed the entire exam. The remaining 17 students (about 3 percent) failed because of the limits placed on compensation between the part scores. Of these 17 students, 11 failed because their essay scores were lower than 48* and 6 failed because their objective scores were lower than 45**. Projected to the total group, it can be estimated that only 1 or 2 percent (i.e. 40 to 80 students) failed because their objectives scores were below 45 even though they earned Composite Scores of 100 or more. An estimated 2 to 6 percent (i.e. 80 to 240 students) failed because their essay scores were lower than 48 even though their Composite Scores were 100 or higher.

IX. Correlations among the Parts.

Correlations among the objective, Raw Essay Total, Converted Essay Total, and Composite Scores were obtained from data available for the Equating Sample. If correlation tables become available for the total group, those tables should replace these but these correlations should be fairly reliable estimates of the corresponding total group data.

The correlation of .547 between the objective and essay part scores seems about ideal for this type of examination. It is high enough to show that the parts are testing related abilities and consequently that combining the part scores to obtain a single composite score is justified. However, the correlation is low enough to justify the use of both parts; that is, each part is contributing some unique component of its own.

The correlation of .999 between the raw and converted essay scores indicates a strong linear relationship between the scores.

---

*Of these eleven students, eight had scores of 46 and three had scores of 41 (i.e. raw essay scores of 12 and 10, respectively.)

**Of these six students, three had scores of 44 and one each had scores of 43, 42, and 40.

Table 3

Correlations among Part Scores and Composite for the Equating Sample

Score	Objective	Raw Essay	Converted	Composite
Objective	1.000	0.547	0.547	0.877*
Raw Essay Total	0.547	1.000	0.999	0.876*
Converted Essay	0.547	0.999	1.000	0.877*
Composite	0.877*	0.876*	0.877*	1.000

*Spuriously high because the part is included in the total.

As noted earlier, a linear equating method could have been used instead of the more complex curvilinear method with essentially the same results.

X. Reliability.

(1) Reliability of the Objective Part. The CLEP Score Interpretation Guide[4] reports that the reliability of the "Analysis and Interpretation of Literature" examination is .88.

(2) Reliability of the Essay Part. Probably the most relevant kind of reliability coefficient for an essay test would be "form-to-form" reliability; that is, the coefficient that would be obtained by constructing two equivalent forms of the essay part (either of which could be used in the actual administration), administering them both to a sample of students, and obtaining the correlation coefficient between the scores on the two forms. Each student in the sample would take both tests. Since this type of data is not available, the reliability of the essay part must be estimated by other methods. Two methods are used; one method gives an upper limit, the other gives a lower limit. We can be reasonably certain the actual reliability lies somewhere between these two values.

(3) The Upper Limit. This method is based on the correlation between grades on the first and second readings of each essay question. The correlation between the grades for the first and second readings of Question 1 is about .66, and that of Question 2 is about .72. These values are the reliabilities for a simple essay grade. That is, each of the grades for the two readings of Question 1 has

4. _____, Score Interpretation Guide, College-Level Examinations Program, College Entrance Examination Board, Princeton, N. J., 1967, page 31.

a reading reliability of .66 and each grade for Question 2 has a reliability of .72. Since the actual final grade includes the grades for both readings, the reading reliability for each question is higher than that of a single reading. The Spearman-Brown Prophecy Formula[5] can be used to estimate the reliability of the total for each question. For Question 1, the reading reliability is about .80 and for Question 2, about .84. The Essay Total score is the sum of grades for all four readings and the estimated reading reliability for the whole essay part is .90.

Since the unreliability due to grading is only one of several sources of unreliability contained in the scores, the reading reliability is always greater than the score reliability and is an upper limit of score reliability. The reliability of the essay part would not be greater than .90.

(4) The Lower Limit. The second method of estimating score reliability is based on the correlation between the scores for Question 1 and Question 2. This is an approximation to the form-to-form reliability, using each question as if it were a one-question test. To the extent that the different questions on one form are designed to measure different abilities within the general domain covered by the test, the questions within one form may not be equivalent in the same sense as two different forms are equivalent. Therefore, the coefficients obtained in this way may be expected to underestimate the actual score reliability and may be regarded as a lower limit of the reliability of the essay section.

Based on the data from the Equating Sample, the correlation between the grades for Questions 1 and 2 is about .41. This is an estimate of the reliability of each "one-question" test. The actual test consists of both questions so the reliability is higher. The Spearman-Brown Formula is used to estimate the reliability of the entire essay part and gives an estimate of about .58. The actual score reliability is probably between .58 and .90.

(5) Reliability of the Composite Scores. Based on reliability coefficients of .88 for the objective part and between .58 and .90 for the essay part, it is estimated that the reliability of the Composite Score is between .83 and .93.

XI. Validity.

The CLEP Score Interpretation Guide[6] reports correlation of .42 and .54 between scores on the "Analysis and Interpretation of Literature" examination and course grades before and after the final

---

5. Op. cit., Educational Measurement, pages 71-72.

6. Op. cit., pages 33-34.

examination, respectively. The final course grades may or may not include the score on the CLEP exam and the grades were given at a variety of colleges with differing grading systems.

Using data from the CSUC Norms Sample, a correlation coefficient of .39 between Essay Total Scores and course grades was obtained. The cautions suggested in the CLEP Guide and the cautions noted in section III (5) above should be considered in interpreting this correlation.

The California State University

ENGLISH EQUIVALENCY TEST

STATISTICAL REPORT

February, 1974 William R. Cowell

I. The Test.

The California State University English Equivalency Test was administered to 4,071 students on May 12, 1973. The test consists of two ninety-minute parts; an objective (multiple-choice) part and an essay part. The objective part is the Subject Examination "Analysis and Interpretation of Literature" of the College-Level Examination Program (CLEP). The essay part consists of two questions prepared and graded by the English professors of the California State University and Colleges (CSUC).

II. Scores.

(1) Objective Score. The Educational Testing Service scored the objective part and reported scores to the students and the CSUC English Test Center on the CLEP score-scale of 20 to 80.

(2) Essay Question Scores. Each of the two essays was graded by two readers who assigned scores on a scale of 1 to 6. Zero was used to denote that the student did not attempt to answer the question.

(3) Raw Essay Total Score. The four essay question scores were added to obtain the raw essay total score. The essay total scores range from 0 to 24.

(4) Converted Essay Score. The raw essay total scores were transformed to the same scale as the objective scores to permit addition of objective and essay scores in a meaningful way. The technique for converting the scores is described in Section IV.

(5) Composite Scores. Composite scores for the whole examination were obtained by adding the Converted Essay Scores and the Objective Scores.

III. Candidate Groups and Norms Samples.

(1) Total Group. The total group consists of the 4,071 students who took the English Equivalency Test in May 1973.

(2) Equating Sample. The equating sample consists of the first 536 papers completed at the reading session. Essay books were scrambled before the reading to essentially randomize the order. The sample is not significantly different from the total group with respect to the mean and standard deviation of either the objective scores or the essay scores.

Table 1

Comparison of Equating Sample and Total Group

	Objective Scores		Essay Scores	
Group	Mean	S.D.	Mean	S.D.
Total Group	47.45	8.85	47.92	8.40
Equating Sample	47.63	8.75	48.25	8.86

(3) CLEP Norms Sample. The CLEP Norms Sample consists of the 541 college students who participated in the 1964 norming administration of the CLEP Analysis and Interpretation of Literature Examination.

(4) Texas Norms Sample. The Texas Norms Sample consists of the 188 students who participated in the 1970 validity study of the CLEP examination at the University of Texas.

(5) CSUC Norms Sample. The essay part of the English Equivalency Test was administered to a sample of CSUC students completing a year of college English in an attempt to link performance on the test to performance in their course (as measured by their grade in the course). It had been planned to use the scores of the CSUC Norms Sample to establish the minimum passing score for the essay part of the examination. Since fewer than half of the expected number of papers were received in time for the reading and because the scores were unreasonably low, it was decided to base the minimum passing scores on other criteria (see Section VI). A summary of the score information for the Norms Sample is shown in Table 2. These data provide some indication of the relationship between college grades and grades on the essay part of the exam.

Table 2

Summary Data for CSUC Norms Sample
Raw Essay Total Scores

Course Grade	N	Mean	Standard Deviation
A	26	13.5	4.0
B	75	12.9	3.0
C	98	10.6	3.5
D	33	10.1	3.2
F	10	7.8	2.4
Total	242	11.5	3.7

IV. Equating the Essay and Objective Scores.

"Equating" scores means making them equivalent with respect to some characteristic. Here, the essay scores were transformed to the same scale (20 to 80) as the objective scores so that the transformed essay scores could be added to the objective scores to obtain a composite score in which the two parts would be equally weighted. Had the raw essay scores been added to the objective scores, the essay part would have had a weight of 1/4 of the total. Also, the transformation was done in such a way that the converted essay scores can be directly compared with objective scores. That is, a converted essay score of 48, for example, represents the same level of performance as an objective score of 48.

Of several acceptable methods of equating which are available,[1] the equipercentile (curvilinear) method was selected because the nature of the distribution of essay scores could not be predicted. [See V (3)]. The converted score corresponding to each of the raw essay scores was found by the procedure illustrated in the following example. To find the converted score corresponding to a raw essay score of 15, for example, we observe from the distribution of essay scores that about 68 percent of the group had scores below 15. We then look down the "percent below" column in the distribution of objective scores for the percent-below closest to 68. The objective score corresponding to that percent-below is designated as the scaled score for an essay score of 15. In this case, the objective score of 53 corresponds to 68 percent below. So a raw essay score of 15 is equivalent to a scaled score of 53. [See V (2)]. Similarily, a scaled score is found for each raw essay score. Table 3 shows the converted score corresponding to each raw essay score.

V. Technical Notes About the Equating.

(1) The equating was done using data from the Equating Sample described in III(2) above. The sample was used instead of the total group so that the statistical work could proceed along with the essay reading; i.e., score conversions and distributions could be prepared for the sample data in time for a meeting with Question Leaders at the reading. The cost of equating was also reduced by using a sample rather than the total group. Scrambling the essay booklets before the reading provided a sample which was representative of the total group.

(2) A method of graphical smoothing, described in "Educational Measurement"[2], was used to interpolate between percentile ranks in the distribution of objective scores.

---

1. Angoff, W. H.; "Scales, Norms, and Equivalent Scores" in R. L. Thorndike (Ed.) Educational Measurement, 2nd Ed., Washington, D. C.: American Council on Education, 1971, pp. 563-590.

2. Ibid., page 572.

(3) Although the equipercentile method was selected because it allows for a non-linear relationship between objective and essay scores, the correlation between raw essay and converted essay scores is .999. This extremely high coefficient of linear relationship indicates that the simpler linear methods[3] of equating could have been used with essentially the same results as the equipercentile method.

Table 3

Score Conversion Table
CSUC English Equivalency Test

May, 1973

Essay Score	Scaled Score
24	76
23	73
22	71
21	69
20	67
19	64
18	62
17	59
16	56
15	53
14	51
13	48
12	46
11	43
10	41
9	38
8	35
7	33
6	31
5	28
4	26
3	24
2	22
1	20
0	18

3. Ibid., page 513.

VI. Minimum Passing Scores.

The minimum passing score for the total test was determined by setting a passing point for each part and adding the scores. For both the CLEP Norms Sample and the Texas Norms Sample, described in paragraphs III (3) and III (4) above, the average CLEP score for "C" students was 49. It was decided that 49 should be the minimum passing level for the objective part. The minimum passing score for the essay part was set at 14, which converts to a scaled score of 51. A composite score of 100 (49 objective + 51 essay) was thus determined to be the minimum passing score on the examination.

Setting a single cut-off point on the Composite Score allows for freely compensating part-scores; that is, a very poor performance on the essay part can be compensated for by a very good performance on the objective part, and vice versa. It was decided that limits should be placed on this compensation. In addition to earning a composite score of 100 or more, a student had to achieve a minimum score on each part. These minima were set at 45 for the objective part and 48 for the essay part. A converted score of 48 corresponds to a raw essay score of 13. A student who scored 12 or lower on the essay part could not pass the exam no matter how high the objective score. A raw essay score of 12 represents the score level of uniform agreement among four readers that the essay performance is unsatisfactory. An objective score of 45 was set as the cut-off point because in both the CLEP Norms Sample and the Texas Norms Sample a score of 45 minimized the number of A, B, and C students who would fail and the number of D and F students who would pass.

VII. Number and Percent Passing.

Of the 4,071 students who took the test, approximately 33 percent of them (1,362 students) passed and are eligible to receive six semester-hours credit. From the distribution of scores on the objective part it can be determined that approximately 43 percent "passed" the objective half of the test - that is, they earned scores of 49 or higher on the objective part. Therefore, had only the objective part of the test been used, and the same minimum passing score established, an additional 10 percent would have passed. These students failed because their essay scores were unsatisfactory.

Approximately 40 percent of the total group achieved "passing" essay scores - that is, they earned scores of 51 or higher on the essay part of the test. Therefore, had only the essay part been used and the same passing point established, an additional 7 percent would have passed.

In addition to earning a total score of 100 or higher, a student had to earn scores of at least 45 on the objective part and 48 on the essay part. Of the total group, 153 students achieved scores of 100 or more on the total test but did not pass because they did not earn the minimum score on one of the parts.

VIII. Reliability.

(1) Reliability of the Objective Part. The CLEP Score Interpretation Guide[4] reports that the reliability of the "Analysis and Interpretation of Literature" examination is .88.

(2) Reliability of the Essay Part. Probably the most relevant kind of reliability coefficient for an essay test would be "form-to-form" reliability; that is, the coefficient that would be obtained by constructing two equivalent forms of the essay part (either of which could be used in the actual administration), administering them both to a sample of students, and obtaining the correlation coefficient between the scores on the two forms. Each student in the sample would take both tests. Since this type of data is not available, the reliability of the essay part must be estimated by other methods. Two methods are used; one method gives an upper limit, the other gives a lower limit. We can be reasonably certain the actual reliability lies somewhere between these two values.

(3) The Upper Limit. This method is based on the correlation between grades on the first and second readings of each essay question. The correlation between the grades for the first and second readings of Question 1 is about .66, and that of Question 2 is about .72. These values are the reliabilities for a single essay grade. That is, each of the grades for the two readings of Question 1 has a reading reliability of .66 and each grade for Question 2 has a reliability of .72. Since the actual final grade includes the grades for both readings, the reading reliability for each question is higher than that of a single reading. The Spearman-Brown Prophecy Formula[5] can be used to estimate the reliability of the total for each question. For Question 1, the reading reliability is about .80 and for Question 2, about .84. The Essay Total score is the sum of grades for all four readings and the estimated reading reliability for the whole essay part is .90.

Since the unreliability due to grading is only one of several sources of unreliability contained in the scores, the reading reliability is always greater than the score reliability and is an upper limit of score reliability. The reliability of the essay part would not be greater than .90.

(4) The Lower Limit. The second method of estimating score reliability is based on the correlation between the scores for Question 1 and Question 2. This is an approximation to the form-to-form reliability, using each question as if it were a one-question test. To the extent that the different questions on one form are designed to measure different abilities within the general domain covered by the test, the questions within one form may not be

---

4. , Score Interpretation Guide, College-Level Examinations Program, College Entrance Examination Board, Princeton, N. J., 1967, page 31.

5. Op. cit., Educational Measurement, pages 71-72.

equivalent in the same sense that two different forms are equivalent. Therefore, the coefficients obtained in this way may be expected to underestimate the actual score reliability and may be regarded as a lower limit of the reliability of the essay section.

Based on the data from the Equating Sample, the correlation between the grades for Questions 1 and 2 is about .41. This is an estimate of the reliability of each "one-question" test. The actual test consists of both questions so the reliability is higher. The Spearman-Brown Formula is used to estimate the reliability of the entire essay part and gives an estimate of about .58.

(5) Reliability of the Composite Scores.[6] Based on a reliability coefficient of .88 for the objective part and a lower limit of .58 and an upper limit of .90 for the essay part, it is estimated that the reliability of the Composite Score is between .83 and .93.

IX. Correlations Among the Parts

Table 4

Correlations Among Part Scores and Composite for the Equating Sample
N = 536

Score	Objective	Raw Essay	Converted	Composite
Objective	1.000	0.547	0.547	0.877*
Raw Essay Total	0.547	1.000	0.999	0.876*
Converted Essay	0.547	0.999	1.000	0.877*
Composite	0.877*	0.876*	0.877*	1.000

*Spuriously high because the part is included in the total.

Correlations among the Objective, Raw Essay Total, Converted Essay Total, and Composite Scores were obtained from data available for the Equating Sample. The correlation between the objective and essay part-scores is 0.547. This correlation coefficient is an index of the extent to which the objective and essay parts measure the same abilities. If both parts were perfectly reliable, the correlation coefficient would be higher; i.e., the correlation coefficient is less than 1.000 partly because the objective and essay tests measure different abilities and partly because they are less than perfect measuring instruments. The correlation

6. Ibid., pages 401-404.

coefficient can be adjusted to eliminate the effect of the unreliability of the tests.[7] Using .88 as the reliability of the objective part and a lower limit of .58 and an upper limit of .90 for the reliability of the essay part, the corrected (disattenuated) correlation between the objective and essay parts is estimated to be between .61 and .77.

The correlation between the objective and essay part-scores seems about ideal for this type of examination. It is high enough to show that the parts are testing related abilities and consequently that combining the part scores to obtain a single composite score is justified. However, the correlation is low enough to justify the use of both parts; that is, each part is contributing some unique component of its own.

The correlation of .999 between the raw and converted essay scores indicates a strong linear relationship between the scores. As noted earlier, a linear equating method could have been used instead of the more complex curvilinear method with essentially the same results.

X. Validity

The CLEP Score Interpretation Guide reports correlations of .42 and .54 between scores on the "Analysis and Interpretation of Literature" examination and course grades before and after the final examination, respectively. The final course grades may or may not include the score on the CLEP exam and the grades were given at a variety of colleges with differing grading systems.

Using data from the CSUC Norms Sample, a correlation coefficient of .39 between Essay Total Scores and course grades was obtained. The cautions suggested in the CLEP Guide and the cautions noted in Section III (5) above should be considered in interpreting this correlation.

XI. Summary

The California Sate University English Equivalency Test was administered to 4,071 students in May 1973. The test consists of two parts; an objective (multiple-choice) part and an essay part. The essay part was also administered to a norming sample of 242 CSUC students. Essay scores were converted to the same scale as the objective section scores by the equipercentile (curvilinear) method. Data suggests that simpler linear methods could have been used with essentially the same results.

To pass the exam, a student had to earn a total score of 100 or higher and, in addition, earn scores of 45 or higher on the objective part and 48 or higher on the essay part. Approximately 33 percent of the students passed the exam.

---

7. Lord, F. M. and Novick, M. R.; Statistical Theories of Mental Test Scores, Reading, Mass.: Addison-Wesley Publishing Company, 1968, pages 115-118.

The correlation between the objective and essay part scores is about .55; high enough to show the parts are related and low enough to show each part contributes some unique component to the total score. The reliability of the objective part is about .88 and that of the essay part is probably between .58 and .90. The reliability of the composite (objective plus essay) is probably between .83 and .93.

The correlation between part-scores and college course grades is approximately .40 for both the objective and essay parts.

PUBLICITY FOR 1973-74 ENGLISH EQUIVALENCY EXAMINATION

Publicity pertaining to the statewide English Equivalency Examination has included both program information and editorial comment. In December 1972, William Trombley in the Los Angeles Times reported on the agreement reached between the Chancellor's Office and the California State University and College English faculties for the English Equivalency Examination, and on the nature, scope and purpose of the proposed examination. During its development in early 1973, the project as well as the concept of English Equivalency Testing came under professional scrutiny in the Chronicle of Higher Education.

As firm timetables were established for the testing phase of the program, news releases from the Chancellor's Office and each of the participating institutions provided notice of test dates, sites and prerequisites for participation. More than seventy-five newspapers throughout the state, including those with local and statewide circulation, carried pretest stories about the project. The pretest publicity campaign was rounded out by radio appearances of the project director, and by direct mailings of test brochures and application forms to all state high school counselors, and to all persons applying at that time for admission to any of the California State Colleges and Universities. In all, approximately 59,000 registration-information forms were distributed.

Media coverage of the project following the May examination focused primarily on the test results. For example, the test outcomes reported by the Chancellor's Office in a news release

dated June 28, 1973, were carried in the Los Angeles Times on the following day under the byline of William Trombley who referred to the success of the 1973 test and the possibilities of refunding the program in 1974. The Los Angeles Herald Examiner carried an editorial on July 4, and many other newspapers throughout the state reported the test results.

Public interest in the program seemed to remain fairly high.

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES

Office of the Chancellor

NEWS RELEASE

5670 WILSHIRE BOULEVARD • LOS ANGELES, CALIFORNIA 90036 • PUBLIC AFFAIRS (213) 939-2981, EXT. 231

73-10 FOR IMMEDIATE RELEASE February 25, 1973

LOS ANGELES--The first systemwide program enabling students to earn academic credits without classroom attendance was announced today by Chancellor Glenn S. Dumke of The California State University and Colleges.

The program will permit seniors in all California high schools to be tested May 12, 1973, on their knowledge of freshman English. Those who pass will receive credit equal to six semester units at any of the system's 19 campuses.

Under the plan, incoming freshmen who demonstrate sufficient college-level ability will be able to move directly into advanced course work in English.

Application forms will be mailed between March 1 and April 18 to students who have already applied for fall 1973 admission. The forms must be returned by April 27, 1973. Forms also will be available from counselors in high schools.

The tests will be administered on all California State University and Colleges campuses only the one day--Saturday, May 12. The campus on which the applicant takes the test need not be the one he plans to attend.

"We are moving rapidly on the premise that many students come to our campuses with sufficient knowledge and ability to proceed directly into more advanced levels of academic work," Chancellor Dumke said.

"The new testing program developed by representatives of our English faculties is an important example of our efforts to increase the options of higher education. It contains the hallmarks of much success."

Dr. Edward M. White, chairman of the Department of English at California State College, San Bernardino, the project's director, said the test is divided into two 90-minute components with a break in between.

The first component is an objective exam, the College-Level Exam-

--MORE--

California State University, San Jose (1857) - California State University, Chico (1887) - California State University, San Diego (1897) - California State University, San Francisco (1899) - California Polytechnic State University, San Luis Obispo (1901) - California State University, Fresno (1911) - California State University, Humboldt (1913) - California State Polytechnic University, Pomona (1938) - California State University, Los Angeles (1947) - California State University, Sacramento (1947) - California State University, Long Beach (1949) - California State College, Bakersfield (1965) - California State University, Fullerton (1957) - California State University, Hayward (1957) - California State College, Stanislaus (1957) - California State College, Dominguez Hills (1960) - California State College, Sonoma (1960) - California State College, San Bernardino (1965) - Colorado State College, Dominguez Hills (1960) - California State College, Bakersfield (1965)

ination Program (CLEP) Subject Examination in Analysis and Interpretation of Literature by the Educational Testing Service of Princeton, N.J.

Sponsored by the College Entrance Examination Board, objective-type CLEP tests were tried on a large-scale exploratory basis in 1971 with freshmen about to enter California State University, San Francisco, and California State College, Bakersfield.

Although CLEP tests in the new project are not the same as those in the 1971 program, experience from the earlier project prompted wide agreement among English faculties that fully adequate testing must contain both objective and written portions.

Thus, the test's second component will be in essay form. It will consist of two questions composed for students by English professors throughout The California State University and Colleges.

Students taking the exam will pay a $15 fee, the standard cost for the CLEP test alone. The system's Fund for Innovation and Improvement will finance the project's administration and the separate essay component.

Results of the CLEP portion of the test will--at the option of the student--be sent to other institutions as well as those in The California State University and Colleges.

However, only the 19 campuses in the State University and Colleges system will be provided a list of students who pass both parts of the exam. These students, upon registration, can claim--and will automatically receive--academic credit.

Passing scores will be based on average performance of "C" students who have actually completed six semester units of freshman English. Students not attaining this level will not be penalized, however, as none of their names will be forwarded to admissions offices.

"We expect 3,000 to 4,000 students will take the examination," Dr. White said. "Many able high school seniors should pass, if they have done enough reading and writing."

Dr. White noted that, although English professors usually oppose mass testing programs, this one has received unanimous approval of English Department chairmen because of its essay component and the high quality of the objective-type CLEP test.

"The entire process is being supervised by English professors," he said. "The professors have full confidence in this project and urge all students who feel they might pass to take advantage of this

--MORE--

innovative opportunity to gain a head start in college."

In addition, the project has the endorsement of the Chancellor's Council of Presidents and the Academic Senate of The California State University and Colleges.

Dr. White, as project director, will be assisted by Dr. Richard Lid, chairman of the California State University, Northridge, English Department. Dr. William Abbott, chief test officer at California State University, Long Beach, will be coordinator of test administration.

Dr. White said students taking the examination will be informed of the results in two stages. They will receive results of the CLEP portion IN June from Princeton, N.J., and the combined results of both CLEP and essay components in July or August.

#

NOTE: For additional information please contact Public Affairs, Office of the Chancellor (213) 938-2981, Ext. 231, or Dr. Edward M. White, CSC, San Bernardino (714) 887-6311, Ext. 597.

Clipping of

Los Angeles Times Sun. Dec. 17, 1972, pg. 173, removed due to poor reproducibility.

THE CALIFORNIA STATE UNIVERSITY AND COLLEGES • OFFICE OF THE CHANCELLOR

NEWS RELEASE

5670 WILSHIRE BOULEVARD, LOS ANGELES CALIFORNIA 90036 • PUBLIC AFFAIRS, ROOM 1215, (213) 938-2981, EXTENSION 231

73-46 FOR IMMEDIATE RELEASE June 28, 1973

LOS ANGELES--More than 1,300 newly graduated high school seniors will receive academic credit for freshman English upon entrance in The California State University and Colleges, Chancellor Glenn S. Dumke announced today.

The seniors are among approximately 4,100 who last month took the California State University English Equivalency Test, the first exam of its kind administered under auspices of the 19-campus system.

By passing the two-part examination, students can receive credit equal to six semester units at any of the system's 14 universities and five colleges.

Thus, they can move directly into advanced coursework in English and--if they choose to do so--shorten the usual class attendance time required for graduation.

The examination program was supported in part by the CSUC Fund for Innovation, created by the Legislature last year in response to Chancellor Dumke's 1971 proposals for New Approaches to Higher Education.

English professors throughout the system devised and conducted the testing plan under the direction of Dr. Edward M. White of California State College, San Bernardino, and Dr. Richard Lid of California State University, Northridge.

--MORE--

"The results are most encouraging in that many students will not be required to duplicate work they have had in high school and can use their time in higher education more effectively," Chancellor Dumke said.

"I am pleased with both the results of the examination and the adherence to established academic principles which was strongly evidenced in this unique testing program."

Dean David H. Provost of the Division of New Program Development and Evaluation, which monitors the system's approximately 50 pilot projects in educational innovation, said students who took the test will be notified of their eligibility to receive academic credit in letters being mailed this week.

Congratulatory letters to the some 1,343 students eligible for academic credit noted that they "have demonstrated the same level of ability many students achieve only after a full year's work in college English."

The test, divided into two 90-minute components, was administered May 12 at all system campuses.

The first component was an objective exam, the College-Level Examination Program (CLEP) Subject Examination in Analysis and Interpretation of Literature. The CLEP portion was developed by the Educational Testing Service of Princeton, N.J., for the College Entrance Examination Board.

Dr. White said California students performed very well on the CLEP component; the test group, without the college course, approached the national norm based on students who have completed the college course.

The test's second component required students to provide essay

responses to two questions designed by English professors from The California State University and Colleges.

Seventy-five English professors graded the essays during a four-day session this month. Essays were read by four faculty members, working independently and without knowledge of grades assigned by the other readers.

Upon completion of the individual grading process, the ratings of all four readers formed the basis for assignment of a final grade.

Combination of scores on the CLEP component and the essay grades resulted in the determination of whether students qualified for academic credit in the State University and Colleges.

Students taking the examination paid a $15 fee, the standard cost for the CLEP test alone. The system's Fund for Innovation financed the project's administration and the essay component.

#

NOTE: For additional information please contact Public Affairs, Office of the Chancellor (213) 938-2981, Ext. 231, or Dr. Edward M. White, CSC, San Bernardino (714) 887-6311, Ext. 597.

Clipping of the Los Angeles Times, Part II
Friday, June 29, 1973, pg. 177, removed due
to poor reproducibility.

College-Level Equivalency Exams in English Draw Fire

By Edward R. Weidlein

The use of standardized tests to allow college freshmen to skip introductory courses is being subjected to increasing scrutiny, particularly by professors of English.

The current debate over credit by examination recalls faculty opposition in the 1950's and early 1960's to the advanced-placement tests of the College Entrance Examination Board.

This time, however, a newer series of College Board tests—the College Level Examination Program (CLEP)—is being looked at.

The CLEP tests were instituted in 1965 by the College Board as a way to provide academic certification for knowledge that people picked up outside of college classrooms.

That use of the test has been heavily promoted. In one CLEP television commercial, a tall, bearded job applicant named Lincoln is turned away by a sleazy employment agent because he does not have a college degree—despite Lincoln's plea that he has read a lot, "sort of on my own."

Avoiding Freshman Requirements

The idea of using CLEP to give academic credit to contemporary Lincolns has been generally applauded, but a growing number of colleges have also begun to use the tests to allow incoming freshmen to avoid normal freshman requirements, such as English composition.

Some history professors, mathematics professors, and a few others have criticized this latter use of CLEP, but English professors have been most vocal and have organized several conferences this year to consider how to respond to the burgeoning practice of credit by examination.

Their concern is hardly surprising. Already the profession is wary, because many colleges have been dropping or reducing freshman-English requirements, thereby at least potentially endangering the jobs of some English professors. Now CLEP has the further potential of enabling huge numbers of freshmen to skip many of the freshman-English courses that still are required.

These fears have combined with skepticism that an objective, multiple-choice test can really measure a student's writing ability, and with distrust of administrators who impose credit-by-examination programs. The concern sur-

Continued on Page 6, Column 3

THE CHRONICLE OF HIGHER EDUCATION
March 12, 1973

College-Level Equivalency Tests in English Held Inadequate by Some Professors

179.

Continued from Page 1

faced in California about a year ago, when the California State University and Colleges, in an experiment, gave a battery of five CLEP tests to entering students at two of its 19 campuses.

Of the 873 students who took the tests at California State University at San Francisco, 331—38 per cent—passed all five and became "instant sophomores." About 10 per cent of the 148 students who took the test at California State College at Bakersfield also scored well enough to skip the freshman year.

Several of the students at Bakersfield were quickly counseled by the English department, which had given its own written placement test, to take remedial English before they could expect to take even the regular freshman course.

Experiences like that, and general faculty objections that the English section of the tests did not do well in measuring students' writing abilities, led to the appointment of a university committee to study CLEP. Last fall, the committee produced a 25-page report, and this spring it will conduct its own experiment with a CLEP test.

Few professors challenge the basic idea of a test like CLEP.

"No one should be asked to repeat work in college that he has mastered; he should receive credit for what he knows and proceed to appropriate levels of learning," says Edward M. White, chairman of the English department at California State College at San Bernardino and author of the report on CLEP.

'A Disaster'

But some professors do insist that some CLEP tests are inadequate. "The use of the present CLEP objective tests as a means of exempting a whole year of freshman English is a disaster," charges James A. Parrish, English department chairman at the University of South Florida.

The CLEP tests have been used on a mass basis elsewhere without creating as much furor as they did in California.

At the University of Utah, about 1,200 students were able to trim a full year off the time it will take them to earn a degree. Although Utah has been probably the largest single user of CLEP, faculty criticism of the tests there has been restrained.

Some of the criticism nationwide reflects little more than a common professorial aversion to standardized tests. Mr. White's report tries to combat such perceptions. He points to research that shows that "one good objective test will correlate more highly with a student's writing ability . . . than will one good essay test."

That is more or less accepted gospel among psychometricians, the people who put objective tests together. Less widely held is a view Mr. White puts forth as almost a mandatory corollary: "The combination of objective items (which measure accurately some skills involved in writing) with an essay (which measures directly, if somewhat less accurately, the writing itself) proved to be more valid than either type of item alone."

Edward M. White: "No one should be asked to repeat work in college that he has mastered."

The statement comes, in fact, from *The Measurement of Writing Ability*, published in 1966 by the Educational Testing Service, the semi-autonomous agency that the College Board contracts with to write, administer, and score CLEP and other tests.

Some E.T.S. officials question, however, whether the benefits of including essays with the objective tests are worth the costs of grading the essays, which Mr. White admits is a "hideously expensive" undertaking.

A Question of Costs

He and his colleagues hope to show that the essays are worth the $6 or more it costs to grade each one.

It is a cost the College Board itself cannot hope to bear. The CLEP program has already consumed $3.1-million in grants from the Carnegie Corporation and at least an equal amount in board funds. The number of people taking the tests is mushrooming, however. There were about 20,000 candidates last year and this year the board expects 55,000.

This month, more than 30,000 application forms for a special "California State University English Equivalency Test" were mailed to California high schools. The test, to be given in May, is essentially an objective CLEP test on the "analysis and interpretation of literature," together with an essay question put together by the state system's English professors.

Mr. White estimates that 4,000 high school students will sign up to take the test. Those who get a combined grade of C or better will receive six semester-units of credits in English at the college and university system's 19 campuses.

More Rigorous Test

Students will pay the normal $15 CLEP fee to take the test. Those fees will go to the College Board. The state system has also allocated $40,000 to cover the grading costs of $6 per test plus other expenses.

Implicit in this plan developed by Mr. White and his fellow English professors is the belief that the new, more rigorous CLEP test will not permit as many as 38 per cent of the students who take it to become "instant sophomores."

"If equivalency becomes one more mechanical device to turn education into processing, we will have done our students and our society a significant disservice, even if we have saved them some cash," Mr. White says.

Although CLEP officials express much greater confidence in their whole range of tests than does Mr. White, they have also taken some of his comments to heart. Albert Serling, CLEP's program director at Educational Testing Service, calls Mr. White's report "about the best thing that has ever been done in the field of English testing."

But CLEP officials may have some difficulty in convincing professors that their jobs are not jeopardized by the tests. For instance, James Ballowe, chairman of the English department at Bradley University, states, "The examinations inevitably lead to the cutting down of staff in large freshman composition courses."